THIS WEEK

Look back in wonder

The launch of the first Sentinel satellite heralds an era in which detailed data on everything from earthquakes to urbanization will be freely available to anyone interested in Earth's future.

The Copernicus Earth-observation programme, whose first Sentinel satellite launched last week, is a rare example of a coherent and ambitious Europe-wide vision backed up with the funds needed to achieve it. It shows that Europe can accomplish great things when its politicians and institutions put their minds to it.

The Copernicus programme — a joint initiative between the European Union and the European Space Agency (ESA) that also involves the intergovernmental meteorological agency EUMETSAT — is remarkable in many ways. By the end of the decade, it aims to combine data from satellites and ground-, air- and sea-based sensors to deliver a dizzying array of imagery and information. Policy-makers, water experts, disaster responders, farmers, fishermen and the interested public all stand to benefit. The satellites will make Earth observation as routine a public service as weather forecasting and satellite navigation, and will also be a boon for science.

Dedicated missions will always be needed to seek specific answers, but routine observations are at the heart of a better understanding of the processes taking place in, on and above Earth. The six families of Sentinel satellites will generate almost real-time data on many of the key components of the Earth system (see page 160). And because the satellites will be routinely replaced, the programme should provide decades of continuous data. The value of such data has been amply demonstrated by the US Landsat satellites, which have delivered a priceless 42-year record of changes in Earth's land surface.

Recognizing Copernicus's scientific potential, the European Commission last year decided to extend the scope of the programme beyond practical applications and include academic scientists as a formal user group. This is a wise move, because the scientific community can help to optimize the systems to make the data more useful for research, and also provide expertise to help the programme evolve.

The European Commission has also made the right call in providing all Sentinel data free of charge. This is essential but was far from a given; the free-market-oriented commission was initially keen to charge for the data, but that would have greatly limited their use. Only after a long battle with scientists and ESA officials did the commission agree to making the data free. Helping to sway the argument were studies showing that making data free could generate a proliferation of downstream commercial applications (some of which are likely to come to a smartphone near you), potentially creating billions of euros in economic benefits and tens of thousands of European jobs.

Like Copernicus, the US Landsat programme has at times been hamstrung by free-market ideology. Landsat images used to cost thousands of dollars before the United States decided to make them free in 2008 to maximize the value of the satellite data. As a result, the number of Landsat images downloaded each year rocketed from 15,000 to more than 4 million by 2013.

Copernicus has not been immune to problems, and budgetary wrangling between the European Commission and member states delayed the launch of Sentinel-1 by three years. But it now has guaranteed funding of €4.3 billion (US$5.9 billion) until 2020. Although short of the €5.8 billion sought, the situation contrasts with that in the United States, where budgetary instability and other problems continue to plague the planning and launch of new missions.

Copernicus highlights the continuing political recognition in Europe that proper stewardship of the planet demands the routine collection of data. It makes sense, too, for different programmes to work more closely together. Coordination between satellite programmes — including Copernicus and Landsat — could bring many benefits: the increased frequency of imaging could, for instance, allow the monitoring of changes that occur on short timescales, such as variations in soil moisture.

> *"The satellites will make Earth observation as routine a public service as weather forecasting."*

Indeed, single large satellites such as Landsat may ultimately become a relic of the past. The trend now — and one vaunted by a US National Research Council report last year on the future of Landsat — is for constellations of smaller, cheaper satellites that provide faster revisits and that minimize the impact of satellite failures.

A new breed of Earth-observation start-up firms are taking the 'constellation' route. The instruments carried by their satellites are often rudimentary compared with those of scientific satellites, but they are good enough for many practical applications and for some scientific purposes. Better Earth observation can only increase awareness of the need for action to protect the planet. There is little point in scientists unravelling the mechanisms of the planet's processes if the knowledge cannot be translated into better management of Earth's environments. ∎

Political science

Russia deserves to be sanctioned, but halting scientific collaboration is not the way to do it.

As European Union foreign ministers gathered in Athens at the weekend to discuss a way out of the Ukrainian crisis, violent pro-Russian activists were occupying local administration buildings and hoisting Russian flags in Donetsk, eastern Ukraine. Analysts fear that the riots might have been deliberately launched to provoke Russian troops to move farther into eastern Ukraine, where a large proportion of the population is ethnic Russian.

Western governments have warned the Kremlin and its increasingly unruly leader, Vladimir Putin, off further military adventures. As we

report on page 162, sanctions imposed on Moscow in response to last month's annexation of the Crimean Peninsula are threatening to affect scientific collaborations. Should the crisis continue to escalate, scientific relations between Russia and the West might suffer in earnest — at least for a while.

This would be unfortunate and almost certainly counter-productive. Moscow's violation of international law cannot be tolerated and Putin must know that there will be serious consequences for his attempt to grasp territory that belongs to another country, no matter what historic and cultural ties exist between the two nations. But science and education — spheres with a unique potential to build trust between nations — should not be used as pawns in the current conflict.

A move by NASA to suspend official contacts with Russian space entities (with the exception of activities involving the International Space Station), for instance, was premature and should be revised. The United States and Russia have collaborated in space for 40 years. Their cooperation survived the Soviet invasion of Afghanistan in 1979, when almost all other official ties between the cold war superpowers had been cut.

When the Soviet Union began to fall apart ten years later, Western support of science in the region effectively kept many Soviet nuclear experts from selling their skills to the highest bidder. And US space activities continue to depend to no small extent on Russian launch vehicles and know-how.

This long and fruitful history of scientific cooperation has proved a solid base for Western relations with Russia, and there can be no doubt that international science has benefited from research carried out in labs from Vladivostok to St Petersburg and Kiev. The political success of scientific and academic exchange programmes launched during and after the cold war — including the North Atlantic Treaty Organization's small but perfectly formed Science for Peace and Security Programme, which shut down its Russian operations last week — goes to show that civilian science is a veritable peacekeeping activity.

Given that, it seems absurd to halt collaborations when the political weather turns bad.

Output of Russian science declined substantially after 1990 — Ukrainian science even more so — but expatriate mathematicians, chemists and physicists have brought many skills and fresh ideas to Western labs. The Crimea crisis, for all the concerns it is raising, is no excuse for ending this successful partnership.

In fact, it could become a reason for Western universities, scholars and think tanks to refocus on a part of the world whose social and political dynamics they have neglected as their interest has shifted to regions such as China and the Arab world. People in the West have been surprised by what is happening in Crimea in part because they know so little about the region and its history. And many Western observers are stunned by Putin's resolute pursuit of power, and the level of support that it receives at home, because even in the scholarly world interest in Russian affairs has been low for years.

> *"The Crimea crisis, for all the concerns it is raising, is no excuse for ending a successful partnership."*

Individual scientists may have strong feelings about the independence of Ukraine and Crimea, and many will be disturbed by the nationalist chauvinism that is palpable in some Russian and Ukrainian circles. Some may therefore choose not to attend conferences and meetings in Russia as long as the crisis is raging. But funders and science policy-makers should take a calmer and more strategic view. It would not help the West, Crimea or Ukraine if Russia's current efforts to strengthen and reorganize its struggling science and higher-education systems were to fail. On the contrary: a society that cannot afford to give its best minds the opportunity to pursue science in a free intellectual environment will be more prone to resort to the nationalistic rhetoric and resentment that will fuel conflict — and not only in Ukraine. As in any crisis, the worst thing one can do is to stop talking. ∎

Copper rewired

Two Nature *papers signal new roles for this ancient metal in catalysis and cancer therapy.*

A bizarre 2009 promotional film from the European Copper Institute in Brussels asks viewers to imagine life without its favourite metal, without actually showing how such a life would be any different. Copper probably does not need much promotion anyway. Ever since the first metal workers of antiquity dug the element from a Cyprus hillside and fashioned it into tools, copper has been in high demand. So high, in fact, that researchers in February raised the prospect that international copper production could peak within a few decades. (The original Cypriot copper mine lives on in the chemical symbol Cu, drawn from the Latin name for the island's metal.)

Nature this week would certainly be poorer without copper. The metal is at the heart of two discoveries reported on our website.

The first describes improvements to the intrinsic catalytic properties of copper that could streamline ethanol production (C. W. Li *et al. Nature* http://dx.doi.org/10.1038/nature13249; 2014). Just as important is the starting material of the chemical reaction that the copper helps along — the greenhouse gas carbon dioxide. In theory, better copper catalysts could offer an efficient method to convert carbon dioxide to liquid, carbon-based fuels.

Imagine life with such a better copper catalyst. Intermittent renewable energy such as wind and solar sources could be used to drive the reaction, and would address two major energy and environmental problems at a stroke — what to do with all the carbon

dioxide we generate, and how to store and transport renewable energy. A News & Views article lays out the roadmap for making it happen (A. M. Appel *Nature* http://dx.doi.org/10.1038/nature13226; 2014).

The research demonstrates that all is not lost when it comes to the fight against climate change. More research will produce better and more efficient technology, yielding some new and some improved versions of what we already have.

The traditional copper-catalysed conversion of carbon dioxide to liquid fuel proceeds through an intermediate of carbon monoxide. Many catalysts can perform the first step, but only copper can mix the carbon monoxide with water to produce the fuel. Yet its promise is largely theoretical: the efficiency and selectivity of the reaction until now have been too low for practical use.

In the latest study, chemists at Stanford University in California show that the catalytic properties of copper can be boosted by starting with copper oxide, which is then reduced back to the base metal. Their enhanced catalyst produces more ethanol than a conventional copper catalyst. They suggest that the difference might be down to tiny cracks introduced to the metal, which give the catalyst more space to work.

In the second study, scientists describe experiments that show copper is required for tumour growth and signalling in some cancers, specifically those with a common mutation in the oncogene *BRAF* (D. C. Brady *et al. Nature* http://dx.doi.org/10.1038/nature13180; 2014). Treatments used to mop up excess copper in the body, already used against copper-accumulation disorders such as Wilson's disease, seem to block the growth of these cancer cells too. The experiments were done in mice and cultured human cells, but suggest that broader use of 'copper chelators' against such cancers could be useful. With apologies to the European Copper Institute, life without (as much) copper for some people could be better, not worse. ∎

↻ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

Society needs more than wonder to respect science

Researchers are well placed to explain concepts, but journalists will bring the critical scrutiny needed to integrate science in society, says **Susan Watts***.*

I don't normally watch football on television, but recently I have been paying attention. What has happened in sports presenting, with former and current players replacing specialist journalists, is creeping into science coverage too.

One television executive put it bluntly to me early this year. "We mainly use scientists as presenters, even if it's not their area of expertise. They have more credibility. A journalist would have to have a really unique selling point for us to use them."

By unique selling point, they meant special access, or a personal link to the story being told. But surely journalists already have a unique selling point — they are journalists?

Footballers talking about what we have seen on the pitch can make for cracking analysis. But such coverage will never expose the uncomfortable side of sport that is away from the screen — the drug-taking or the match-fixing. And what about the awkward unease among former heroes-cum-presenters when confronted with the prospect of throwing a tricky question at a current star, even though viewers might be shouting at the TV for them to ask it (my sons among them)?

Sport is not a life-or-death issue, for most people at least. But science and engineering can be. Scrutiny is crucial.

There is a fundamental difference between science communication and science journalism. At the science communication end of the spectrum sit the stories that show people how exciting science can be, the discovery of a wonder material, perhaps, or a new subatomic particle. Explaining the significance of sightings of the Higgs boson or of gravitational waves from the early Universe takes real skill.

Science journalism's job is to tell the stories that explore the murky underbelly of science, like the selling of bogus stem-cell cures to vulnerable patients. It is science journalism that will expose the rushed policy-making, the undisclosed profiteering, the conflicts of interest and the vested interests, the bad experiments, or the out-and-out frauds.

For both, you need to be the kind of person who asks "why" a lot. You need to enjoy coaxing sometimes shy, or reluctant, or just plain difficult scientists to tell you about their work — and then to feel enthused enough to want to tell somebody else.

But a journalist also needs to be persistent, and brave enough to find out the things that people don't want the world to know, and who often work hard to stop the world knowing — and to tell those tales too.

More uncritical wide-eyed stories about the 'wonder' of science at the expense of science journalism is a decadence we can't afford,

WE NEED SCIENCE JOURNALISM TO WEIGH UP THE VALUES AND VICES OF SCIENCE.

intellectually or practically. I am as awestruck as anyone by the beauty of the aurora borealis, but I also want to know more about issues such as what is being called the replication crisis in science. When important cancer papers, for example, can't be reproduced by other scientists, something is wrong. That is hard journalism to do, and even harder to put on television.

So what is behind the subtle shift to science communication, and away from science journalism? One enduring problem is that our media remains dominated by people who are all too often educated in the humanities. Irrespective of how talented these people may be, when the majority of the most-influential roles are filled by people who have no understanding of how science works, then it rarely occurs to them that science is populated by people every bit as interesting and as human as those in the arts or politics, or that the internal battles of science can be every bit as personal and as bitter as any in industry or business.

In broadcast journalism at least, science journalists are increasingly viewed as dispensable as long as a programme editor can dig one out when a big health story surfaces, or in times of extreme weather. If, instead, these editors valued the input of journalists who have a science specialism in the newsroom every day, they would gain not only an eye on issues coming over the horizon, but also the day-to-day drip feed of a scientific perspective into all the stories that appear on a programme. That way, the scientific viewpoint becomes part of a programme's lifeblood, as it should be in a healthy, modern society, and not an added extra.

We need science journalism to weigh up the values and the vices of new science. Without it, we will struggle to place science in its social context as we grapple with its challenges. Take genomic medicine, for example, where there is a balance to be struck between promises of better, personalized health-care and threats to the privacy of our personal data.

The risk is that in our intoxication with the 'wonder' of science, we miss its murkiness. Or worse, we deliberately avoid asking the questions that challenge scientists and technologists about the work they do. Lose that critical perspective, and we lose the ability to take an informed view of what it is we want from science. And do we really want science coverage to vie with that witless brand of sports commentating: "He'll be disappointed with that, Brian"? Indeed, won't we all. ∎

Susan Watts *was science editor of the BBC's* Newsnight *programme until the post was closed in November. This is an edited extract of a talk she gave at last month's Cambridge Science Festival, UK.
e-mail: susan.watts@susanwatts.org*

MICROBIAL GENOMICS

Sequencing spots killer microbes

Sequencing the genome of an antibiotic-resistant microbe can identify the most dangerous isolates, reports a team led by Ruth Massey at the University of Bath, UK.

Her group studied 90 isolates of methicillin-resistant *Staphylococcus aureus* (MRSA) that had varying levels of ability to stick to and kill cultured human cells. The researchers sequenced the isolates, and conducted a genome-wide association study to identify 121 genetic changes, including additions or deletions of DNA, that were linked to this toxicity.

A set of 50 of these variants could predict the most and least toxic of 30 of the isolates, although it was unable to predict the medium-toxicity isolates with any accuracy.
Genome Res. http://dx.doi.org/10.1101/gr.165415.113 (2014)

PLANT BIOLOGY

How bacteria turn plants into zombies

Researchers have uncovered how certain bacterial pathogens that infect plants

EVOLUTION

Hummingbird species on the rise

In just 22 million years or so, hummingbirds have rapidly diversified from a single ancestor into more than 300 species, and some lineages are still generating new species at an extraordinary rate.

Jimmy McGuire at the University of California, Berkeley, and his colleagues compared the DNA sequences of 284 hummingbird species, including *Selasphorus flammula* (pictured).

They found that the birds first diverged from their sister group, the swifts, around 42 million years ago and have diversified into 9 major lineages in South America over the past 22 million years. Speciation has been particularly dramatic in the Andes mountains, which have a wide range of habitats and climates and are home to 40% of hummingbird species.

The researchers calculate that there could be as many as 767 species of hummingbird in the next several million years — more than twice the number that currently exist.
Curr. Biol. http://doi.org/r6v (2014)

make them sterile and capable only of spreading disease.

Phytoplasma pathogens are transmitted by sap-feeding insects that turn flowers (such as *Arabidopsis thaliana*, **pictured** top) into leaf-like structures (bottom) that do not produce seeds. Saskia Hogenhout at the John Innes Centre in Norwich, UK, and her colleagues studied *Arabidopsis* plants and found that a phytoplasma protein, SAP54, interacts with a class of plant proteins called RAD23 to degrade molecules that regulate floral development.

This interaction also seems to boost the attractiveness of infected plants to leafhopper insects, which spread phytoplasma from one plant to another.
PLoS Biol. 12, e1001835 (2014)

MOLECULAR BIOLOGY

DNA regulator acts on RNA too

A DNA-binding protein that regulates several genes also attaches to RNA, revealing another way in which the protein acts as a 'master weaver' of the genome.

Félix Recillas-Targa of the National Autonomous University of Mexico in Mexico City, Danny Reinberg of New York University's Langone Medical Center and their colleagues discovered a region in the DNA-binding protein CTCF that binds to the RNA molecule Wrap53. This RNA regulates the tumour suppressor p53, a protein involved in DNA repair.

When the team mutated CTCF in human cells, CTCF could not bind to Wrap53 RNA and cells failed to trigger responses to damaged DNA, showing that CTCF controls p53 by binding to Wrap53. In a genome-wide screen, the team found CTCF attached to some 17,000 other RNAs.

CTCF could regulate many genes in this way, and binding to RNA could allow the protein to form short chains, which might influence how it creates loops in protein–DNA structures called chromatin, the authors suggest.
Genes Dev. **28**, 723–734 (2014)

PLANETARY SCIENCE

A moon of Saturn hides an ocean

Beneath Enceladus's south pole lies a watery ocean that could hold organic molecules that form the basis of life.

Luciano Iess at the Sapienza University of Rome and his colleagues analysed gravity measurements from the Cassini spacecraft during three flybys of this moon of Saturn in 2010–12. They found more mass than expected at the south pole, suggesting that something denser than ice was beneath the 30–40 kilometres of ice covering Enceladus's surface. The researchers think that a 10-km-thick layer of ocean covers the moon's rocky core.

In 2005, Cassini spotted water vapour and ice spraying through cracks in the icy shell near Enceladus's south pole. These jets could be carrying organic molecules and salts from this ocean to the surface, making the moon a place to search for possible extraterrestrial life.
Science **344**, 78–80 (2014)

CLIMATE SCIENCE

El Niño comes in many flavours

Different kinds of El Niño warming events in the tropical Pacific Ocean can have widely varying effects on global temperatures.

Simon Donner and Sandra Banholzer of the University of British Columbia in Vancouver, Canada, used historical sea surface temperature data to classify El Niño events into three groups on the basis of the location of ocean warming: the eastern Pacific, the central

Pacific and a mixture of both. They then compared their results with three data sets of global temperatures. Although El Niños are commonly associated with warming, the duo found that only the more powerful eastern Pacific El Niño was linked to higher global temperatures.

These varying El Niño effects could explain some of the temperature trends seen in the Pacific Ocean since the late 1880s, such as periods of both accelerated and slow warming.
Geophys. Res. Lett. http://doi.org/r6z (2014)

BIOTECHNOLOGY

Altered trees make digestible wood

Genetically engineered poplars can make a modified polymer in their wood that breaks down more easily than natural forms. Such trees could one day be sources of biofuels.

Current methods for digesting the tough, resistant polymer called lignin involve concentrated chemicals at high temperatures. John Ralph at the University of Wisconsin–Madison and his colleagues inserted a gene into poplar trees (*Populus* spp.) that produces a molecule that is incorporated in the lignin polymer chain. The molecule forms an ester chemical bond in the growing chain, which can be broken down by mild chemicals at lower temperatures, aiding bioethanol production later.

The modified poplars produce normal amounts of lignin and seem to be healthy, but they have not yet been tested in the field.
Science **344**, 90–93 (2014)

AGRICULTURE

Rising carbon saps wheat quality

A field experiment shows that wheat exposed to higher carbon dioxide levels converts less of the nitrogen it absorbs into protein. This could result in decreased

food quality this century.

Arnold Bloom at the University of California, Davis, and his colleagues compared nitrate concentrations and the relative amounts of heavy and light nitrogen isotopes in leaves from wheat (*Triticum aestivum* L.) grown under elevated and normal CO_2 conditions. They found that increased CO_2 slows down the assimilation of nitrate into proteins in leaves.

The predicted rise in CO_2 over the next few decades could decrease protein levels in crop plants by about 3% without sophisticated new approaches to nitrogen fertilization, the authors say.
Nature Clim. Change http://doi.org/r7j (2014)

NANOTECHNOLOGY

DNA robots work in a live cockroach

Molecular 'robots' can perform complex logic tasks inside a living organism.

Ido Bachelet of Bar-Ilan University in Ramat Gan, Israel, and his colleagues used folded strands of DNA to create a suite of nanorobots that open,

CLIMATE CHANGE

European hotspot in a warmer world

★ HIGHLY READ on iopscience.iop.org in March

Even if the global temperature rise is held to a limit of 2 °C, most of Europe will probably warm by substantially more.

A team led by Robert Vautard at the Pierre-Simon Laplace Institute in Gif-sur-Yvette, France, combined 6 global climate simulations with 15 simulations that focused on Europe. The team found that, in winter, warming in northern and eastern Europe would be more intense — with temperatures rising by up to 3 °C — whereas southern Europe would become much hotter in summer.

Winter precipitation would increase over central and northern Europe, and there would be less summer rain in the central and southern parts of the continent, the authors found.
Environ. Res. Lett. **9**, 034006 (2014)

close and coordinate with each other in response to various interactions with certain proteins. When the robots were injected into a living cockroach (*Blaberus discoidalis*; **pictured**), different robot combinations created seven kinds of logic gate that each delivers a different outcome — such as releasing various antibody payloads — on the basis of specific protein cues.

The authors suggest that the technique might eventually be used to control drug delivery in humans.
Nature Nanotechnol. http://dx.doi.org/10.1038/nnano.2014.58 (2014)

↻ **NATURE.COM**
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

DALY AND NEWTON/IMAGE BANK/GETTY

RESEARCH

Distant worlds

Astronomers last week discovered two potential dwarf planets in the outer reaches of the Solar System, orbiting the Sun beyond Neptune. One object, 2013 FY27, is the ninth-brightest trans-Neptunian object known, and at roughly 760 kilometres across is among the largest objects in the outer Solar System. The other object, 2013 FZ27, is smaller and dimmer. The discoveries were made by the same team that last month announced that it had found the Solar System's most distant probable dwarf planet. See go.nature.com/xqiucr for more.

Heavy galaxy

The largest known galaxy cluster in the distant Universe, nicknamed El Gordo — Spanish for 'the fat one' — is 43% heavier than previously thought. The galaxy, officially called ACT-CL J0102-4915, is 3,000 times the mass of the Milky Way. Astronomers think that El Gordo is so big because it formed when two galaxies collided and merged. The estimate comes from measurements taken with NASA's Hubble Space Telescope, which places the galaxy at around 3 billion parsecs from Earth.

POLICY

Russian frost

Both NASA and the North Atlantic Treaty Organization (NATO) cut ties with Russia last week, after it annexed Ukraine's Crimean peninsula. NATO suspended all civilian and military cooperation — including scientific work such as technology on detecting hidden bombs under the Science for Peace and Security Programme. The US
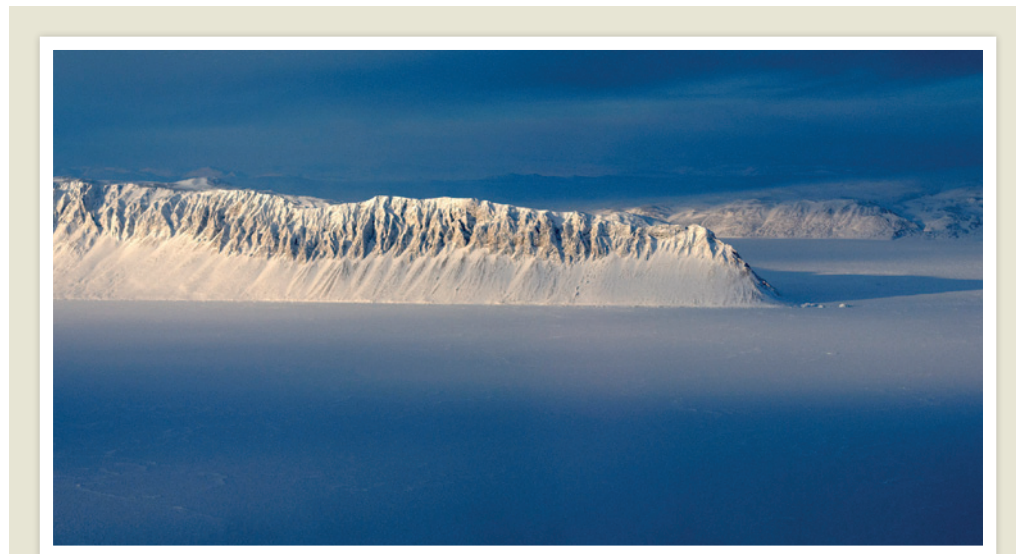
Arctic ice coverage peaks

Arctic sea ice reached its maximum coverage for the year on 21 March, stretching across 14.91 million square kilometres, the US National Snow and Ice Data Center in Boulder, Colorado, said on 2 April. This year's coverage is the fifth-lowest recorded since satellite monitoring began in 1978. Arctic sea ice has declined steadily in recent years, with a record annual minimum in 2012 of 3.41 million square kilometres, 49% below the 1979–2000 average.

government suspended all visits, meetings, e-mails and other contact between NASA and Russian space agencies, except those involving the International Space Station. See page 162 for more.

Culls curbed

The UK government said on 3 April that it will reform its controversial strategy for managing the spread of bovine tuberculosis (TB) by badgers (*Meles meles*). The government's volte-face is effectively an admission that it was wrong to ignore scientists who warned that culling badgers would be ineffective in stemming the spread of bovine TB to cattle. The government said that its pilot programme of badger culls needs to become more effective and humane. Plans to extend the culls to other areas have been postponed. Some UK farmers had argued for culls to protect their herds (see *Nature* **490**, 317–318; 2012).

Whale hunt axed

Japan cancelled its annual Antarctic whale hunt on 3 April in compliance with a 31 March decision from the International Court of Justice in The Hague, the Netherlands, which ruled that the hunt was not scientific (see *Nature* **508,** 12–13; 2014). The hunt was due to start later this year. Japan maintains that its whaling programme is for research and says that it will continue hunting in other areas, including the northern Pacific. Commercial whaling is effectively banned under international law.

Metrics review

The UK government will review how metrics such as citation counts and numbers of published articles are used to assess research quality and award funding to UK universities. James Wilsdon, a science-policy researcher at the University of Sussex, UK, will chair the review panel, which will include *Nature*'s editor-in-chief, Philip Campbell. Wilsdon said that safeguards and guidelines might be needed to avoid the misuse of bibliometrics and the manipulation of the research-assessment system. The review will run until spring 2015.

Clinical trials vote

The European Parliament last week agreed on new rules that could save researchers €800 million (US$1.1 billion) each year in clinical-trial regulatory costs. The rules simplify the submission and assessment of clinical trials, and will make it easier to

conduct large, multinational studies, which could lead to improvements in reliability. Results will also be accessible in a public database to make trials more transparent. The regulation will be implemented in 2016 pending approval by European Union member states.

EVENTS

Atomic timekeeping

US scientists have unveiled the world's most accurate atomic clock. On 3 April, the National Institute of Standards and Technology in Gaithersburg, Maryland, launched NIST-F2, a caesium-based clock that will neither gain nor lose a second in 300 million years, making it about three times more accurate than its predecessor. NIST-F2 will be used to time-stamp US financial transactions, among other applications. Clocks based on caesium are approaching their performance limit; later clocks will probably use atoms that vibrate at higher frequencies, making them more accurate.

Chilean quake

Chile was struck by a magnitude-8.2 earthquake on 2 April that killed at least six people in the north of the country (**pictured**). The quake hit about 100 kilometres northwest of the city of Iquique on Chile's Pacific coast,

according to the US Geological Survey. Despite the region's proximity to a major active fault along the Nazca tectonic plate, the last comparably large tremor to occur in the area was in 1877 — a long period of inactivity called the Iquique seismic gap. See go.nature. com/p22nij for more.

Waste inspection

Employees re-entered the US Waste Isolation Pilot Plant (WIPP) on 2 April for the first time since a leak forced the underground nuclear-waste facility to close in mid-February. Workers found no radioactive contamination during an inspection of the immediate area around the mineshaft entry. They also installed air-monitoring equipment, according to the US Department of Energy. Next week, WIPP workers will begin a search for the source of the contamination at the facility, which is located near Carlsbad, New Mexico.

PEOPLE

AIDS-plan head

The US President's Emergency Plan for AIDS Relief (PEPFAR) has a new leader. Immunologist Deborah Birx took charge of the government's overseas AIDS efforts on 4 April. Birx previously led the Division of Global HIV/AIDS at the Centers for Disease Control and Prevention in Atlanta, Georgia, and oversaw a major HIV-vaccine trial in Thailand during service in the US Army. PEPFAR is the world's largest individual fund to tackle a single disease, having spent more than US$52 billion since 2004.

FACILITIES

Aid-lab initiative

The US Agency for International Development announced on 3 April the launch of a US$100-million Global Development Lab based in Washington DC. The lab will create a pipeline for research and development into food security and nutrition, maternal and child survival, energy access and sustainable water solutions. The initiative marks a shift in the agency's approach to development funding: from helping organizations to tackle specific problems using existing technologies

to identifying problems and funding the development of technological solutions. See go.nature.com/kngsjn for more.

BUSINESS

Pollution payout

Energy company Anadarko Petroleum agreed on 3 April to pay more than US$5 billion to clean up environmental pollution at more than 2,000 sites across the United States. It is the largest ever US settlement for environmental contamination, and brings an end to the lengthy case against energy and chemical company Kerr–McGee, which was bought by Anadarko in 2006. Kerr–McGee was blamed for using cancer-causing coal-tar creosote for wood preservation, and for contamination of soil and groundwater with ammonium perchlorate, a component of jet and rocket fuel.
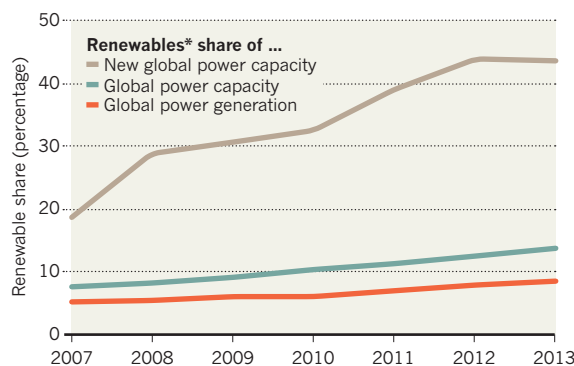
↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news

TREND WATCH

Renewable energy met 8.5% of global electricity needs in 2013, up from 7.8% the previous year (excluding large hydroelectric installations), according to figures released on 7 April by the United Nations Environment Programme. Policy uncertainty and lower solar-electricity installation costs saw total investment in renewable energy fall. But for the first time, China's investment (US$56.3 billion, down 6% from 2012) outweighed that of Europe ($48.4 billion, down 44%).

RENEWABLE RISE
Renewable power continues to climb as a share of total power generation.

Renewables* share of ...
— New global power capacity
— Global power capacity
— Global power generation

Renewable share (percentage)
50
40
30
20
10
0
2007 2008 2009 2010 2011 2012 2013

*Excludes large hydroelectric installations.

NEWS IN FOCUS

Stem-cell biologist Mahendra Rao expected five projects to receive support to set up clinical trials.

REGENERATIVE MEDICINE

NIH stem–cell programme closes

Director resigns as just one clinical–trial award is made.

BY SARA REARDON

Stem-cell researchers at the US National Institutes of Health (NIH) have been left frustrated and confused following the demise of the agency's Center for Regenerative Medicine (CRM). The intramural programme's director, stem-cell biologist Mahendra Rao, left the NIH, in Bethesda, Maryland, on 28 March, and the centre's website was taken down on 4 April. Although no official announcement had been made at the time *Nature* went to press, NIH officials say that they are rethinking how they will conduct in-house stem-cell research.

Researchers affiliated with the centre say that they have been left in the dark. When contacted by *Nature* on 7 April, George Daley, a stem-cell biologist at Harvard Medical School in Boston, Massachusetts, and a member of the centre's external advisory board, said that he had not yet been told of Rao's departure or the centre's closure.

The CRM was established in 2010 to centralize the NIH's stem-cell programme. Its goal was to develop useful therapies from induced pluripotent stem (iPS) cells — adult cells that have been converted into embryonic-like stem cells — and shepherd them towards clinical trials and regulatory approval. Its budget was intended to be $52 million over seven years.

Rao took the helm in 2011. Relations seem to have soured last month owing to an NIH decision to award funding to only one project aiming to move iPS cells into a clinical trial. Rao says he resigned after this became clear. He says that he had hoped that five trials would

be funded, especially because the centre had already sorted out complex issues relating to tissue sources, patents and informed consent.

James Anderson, director of the NIH's Division of Program Coordination, Planning, and Strategic Initiatives, which administered the CRM, counters that only one application — that made by Kapil Bharti of the National Eye Institute in Bethesda and his colleagues — received a high enough score from an external review board to justify continued funding. The team aims to use iPS cells to treat age-related macular degeneration of the retina, and hopes to commence human trials within a few years. Several other proposals, which involved the treatment of cardiac disease, cancer and Parkinson's disease, will not receive funding to ready them for clinical trials. Anderson stresses that Bharti's trial will not be affected by the CRM's closure.

Other human iPS-cell trials are further along. For example, one on macular degeneration designed by Masayo Takahashi at the RIKEN Center for Developmental Biology in Kobe, Japan, began recruiting patients last August.

Anderson says that the CRM will not continue in its current form. "The field is moving so fast that we need to rethink." To that end, the NIH plans to hold a workshop in May to gather stem-cell researchers together and decide what to do with the programme and its remaining budget. "To me that's just smart science," he says. "If something's not on track you don't keep spending money on it."

One option could be to allow CRM projects to be absorbed by the National Center for Advancing Translational Sciences, an NIH institute established in 2011 to translate basic research into therapies. But Anderson says that participants at the workshop will also discuss whether the NIH needs to replace the CRM with another dedicated stem-cell programme.

Rao says he is not yet sure what his plans are. He has already joined the boards of two stem-cell-therapy companies, and eventually hopes to find a company that will allow him to proceed with clinical trials of iPS-cell treatments for Parkinson's disease.

He says that despite what has happened, he is proud of the CRM's accomplishments: its researchers have generated more than 400 lines of iPS cells, and the centre has funded two contractors to create 12 lines of clinical-grade iPS cells that will be available later this year. But, he adds, "It's a little disappointing. The NIH is at a stage where it can be a world leader." ■
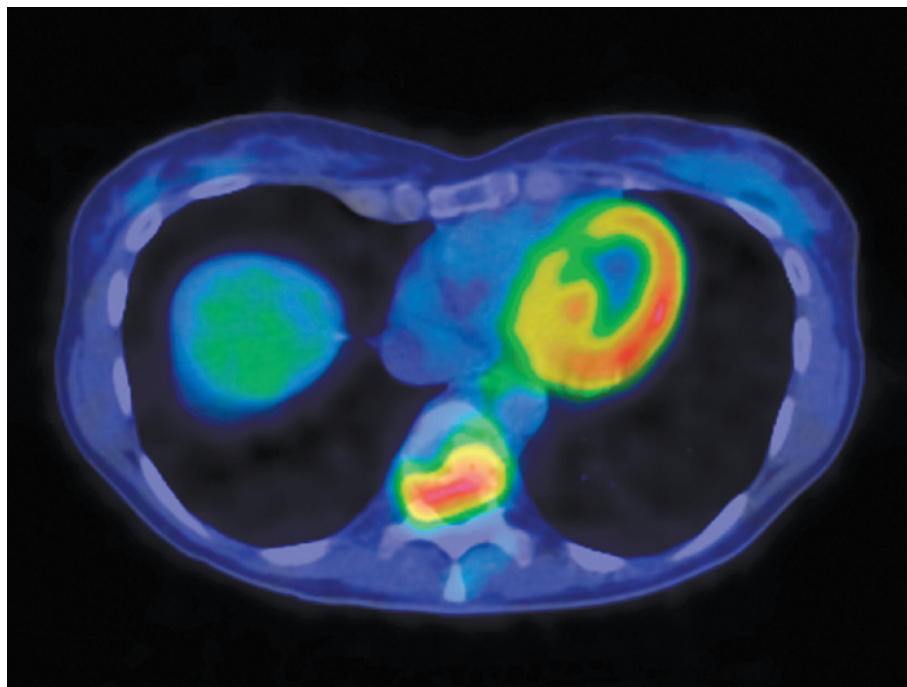
Preferential uptake of glucose by cancer cells allows tumours, here shown in the spine, to be imaged.

CANCER

Metabolic quirks yield tumour hope

Early clinical-trial results show promise for targeting cancer-related biochemical pathways.

BY HEIDI LEDFORD

One of cancer's deadly sins — gluttony — is turning out to be one of its key vulnerabilities.

Cancer cells harness unusual metabolic pathways to obtain the energy and molecular building blocks that they need for their relentless proliferation. Many potential drugs have tried to take advantage of this hunger. Early results for a genetically targeted drug, unveiled this week at the annual meeting of the American Association for Cancer Research in San Diego, California, suggest that the approach could pay off.

"The field is at a turning point," says Almut Schulze, a cancer researcher at the University of Würzburg in Germany.

In some ways, the findings send cancer research back to its roots. For much of the twentieth century, the disease was considered a metabolic malady — an idea that arose in the 1920s, when the German biochemist Otto Warburg showed that cancer cells have

an outsized appetite for glucose. The glucose is broken down, yielding energy in the form of ATP, produced in the cell's mitochondria, as well as components of amino acids, lipids and other compounds needed to build new cells.

The phenomenon, called the Warburg effect, quickly led to techniques that image tumours by tracking radioactively labelled glucose molecules taken up by cancer cells. It also suggested a path to new therapies, says Nissim Hay, a cancer biologist at the University of Illinois in Chicago: "If we can selectively detect cancer cells, why can't we selectively target them?"

In the 1970s, the discovery of chromosomal abnormalities and cancer-causing mutations shifted focus to the disease's genetic origins. Cancer's odd metabolism was ruled an effect, rather than a cause, and researchers largely

"It's very interesting biology that's been neglected for over 30 years."

sidelined its study. But over the past decade, they have come to realize that the mutations alter a handful of key metabolic systems from which cancer derives its energy. These metabolic pathways are potential targets for drugs. "That was the wake-up call," says Matthew Vander Heiden, a cancer researcher at the Massachusetts Institute of Technology in Cambridge.

The clinical work announced at the meeting stems from two targets that were characterized in 2009 by researchers at Cambridge-based Agios Pharmaceuticals. The team reported on the effect of mutations in genes called *IDH1* and *IDH2* that had already been associated with some forms of leukaemia and brain cancer.

The genes encode enzymes that act in an energy-producing metabolic pathway called the citric acid cycle. The Agios team found that the mutations lead to unusually high production of a cancer-promoting compound called 2-hydroxyglutarate (L. Dang *et al. Nature* **462**, 739–744; 2009). This compound, which is produced at only low levels in normal cells, allows cancer cells to proliferate by keeping them in an immature state.

TREATMENT RESPONSE

On 6 April, Agios reported the first clinical-trial results for a drug designed to inhibit the mutant IDH2 enzyme. The trial was small, involving ten patients who had acute myeloid leukaemia with *IDH2* mutations. Three of the patients were removed from the trial after developing infections, a common consequence of advanced leukaemia. But of the remaining seven patients, six responded to treatment. The drug, called AG-221, eradicated cancer cells in five of them.

The results are encouraging, says John Byrd, director of haematology at the Ohio State University Wexner Medical Center in Columbus, who was not involved in the study. The patients in the trial were very sick, he notes, and had already undergone multiple therapies. "This is very impressive," he says.
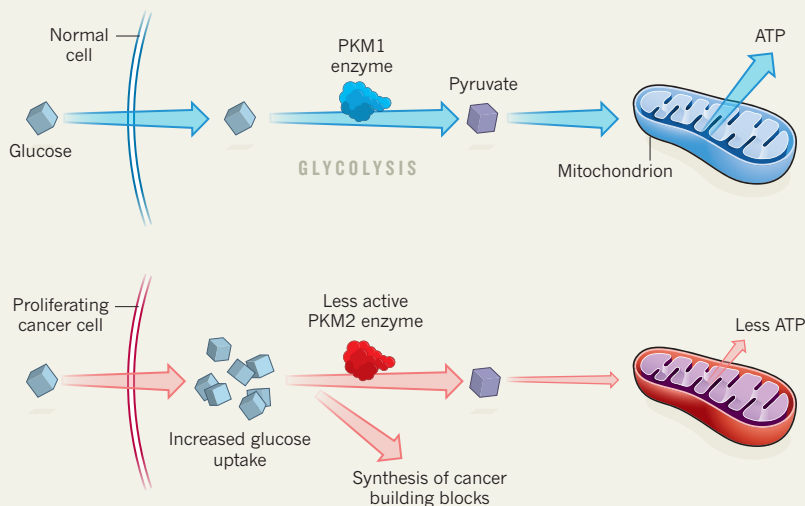
The patients have been on the drug for five months or less, and so it is too early to say how long the effects will last, says Agios chief executive David Schenkein. Tumours quickly become resistant to many drugs that target a single cancer mutation, and metabolic pathways are remarkably plastic, notes Navdeep Chandel, a mitochondrial biochemist at Northwestern University in Chicago. "If a cell needs certain metabolites to grow, it will find a way."

Other drugs based on the quirky metabolism of cancer cells are making their way into the clinic. In February, Calithera BioSciences of South San Francisco, California, launched two trials of a drug that capitalizes on the penchant of some cancer cells to mop up huge amounts of the amino acid glutamine, which they use to make proteins and to fuel the citric acid cycle. The company has produced a

ENERGY HUNGRY

Researchers are trying to find targets within cancer's metabolic cycles, such as one that breaks down glucose.

PUBLIC HEALTH

E–cigarettes affect cells

Questions raised over health effects of devices.

BY DANIEL CRESSEY

Electronic cigarettes can change gene expression in a similar way to tobacco, according to one of the first studies to investigate the biological effects of the devices.

Presented at the American Association for Cancer Research annual meeting on 6 April in San Diego, California, the research looked at human bronchial cells that contained some mutations found in smokers at risk of lung cancer. The cells were immortalized, grown in culture medium that had been exposed to e-cigarette vapour and their gene expression profiled.

The researchers found that the cells grown in medium exposed to the vapour of e-cigarettes showed a similar pattern of gene expression to those grown in a medium exposed to tobacco smoke (S. J. Park *et al. Clin. Cancer Res.* **20,** B16; 2014).

The changes are not identical, says study researcher Avrum Spira, who works on genomics and lung cancer at Boston University in Massachusetts. But "there are some striking similarities", he says. The team is now evaluating whether the alterations mean that cells behave more like cancer cells in culture.

The work is at a very early stage and therefore cannot establish that e-cigarettes can cause cancer *in vitro*, let alone *in vivo*. "They may be safer [than tobacco], but our preliminary studies suggest that they may not be benign," says Spira.

E-cigarettes are extremely controversial. Because they vaporize liquid containing nicotine, rather than burning tobacco, some researchers believe that the devices could greatly reduce the damage done to health by smoking; others, however, argue that they are simply 'renormalizing' smoking. ∎

drug that blocks glutaminase, the enzyme that converts glutamine to glutamate, a key step in the cycle. The hope, says Calithera chief executive Susan Molineaux, is that because of cancer cells' strong dependence on glutamine, concentrations of the inhibitor can be found that will halt cancerous growth without damaging healthy cells.

But finding that sweet spot poses a challenge, because the metabolic oddities associated with cancer are sometimes found in normal cells that need to proliferate rapidly, such as immune cells. Agios was able to sidestep that hurdle by targeting an abnormal form of the IDH2 enzyme associated with cancer.

ENZYME CONUNDRUM

Other approaches have not had the benefit of such a clear target, and have suffered as a result. Several companies, including Agios, have studied an enzyme called pyruvate kinase that helps to break down glucose in the Warburg effect. Most normal cells produce a form of the enzyme called PKM1, but cancer cells tend to favour a less active form, PKM2. That pattern suggested that a drug blocking PKM2 could halt cancer growth.

But Vander Heiden and his colleagues found that tumours grew faster, not more slowly, in mice that lacked PKM2 (W. J. Israelsen *et al. Cell* **155,** 397–409; 2013). To explain this strange result, the researchers hypothesize that multiplying cancer cells may prefer the less active PKM2 — or even no pyruvate kinase at all — because it allows glucose breakdown products to be shunted into pathways that generate molecules needed to build new cancer cells (see 'Energy hungry'). Although that would result in less ATP generated per molecule of glucose broken down, the cells can compensate by taking up more glucose. Nevertheless, the results cast doubt on PKM2 as a drug target. "Initially, people were very excited about PKM2," says Hay. "But then it turned out to be more complicated than expected."

Although Chandel is glad that cancer-metabolism research is taking centre stage, he thinks that expectations for drugs based on the research might be too high — most attempts will not succeed. Even so, he adds, the fervour should lead to a better understanding of metabolic processes in both healthy and diseased cells. "These are good days for studying metabolism," he adds. "It's very interesting biology that's been neglected for over 30 years." ∎

NASA/JPL-CALTECH

→ **MORE ONLINE**

TOP NEWS

Saturn's moon Enceladus hides huge water ocean under its south pole **go.nature. com/vijdmf**

MORE NEWS
- Public disclosure leads to more retractions of flawed publications **go.nature.com/rwos3y**
- Europe urged to relax text-mining access **go.nature.com/gpdekt**
- How Earth got its tectonic plates **go.nature.com/715d6y**

NATURE PODCAST

Flooding the Colorado River; being an IPCC contributor; and amoebae eating **nature.com/nature/podcast**

WATCHERS IN THE SKIES

Once its Sentinel satellites are fully operational, Copernicus will probably be the world's most comprehensive Earth-observation programme. In addition to Sentinels 1 to 6, a Sentinel 5 Precursor satellite will be launched in 2016 to minimize the gap in atmospheric data following the expiration of Envisat in 2012.

SENTINEL-1
Launch date: 1A: Launched; 1B: 2015
Payload: All-weather radar
Revisit time: 1–3 days
Applications: Monitoring sea ice and the Arctic, land surface motion risks, disaster response

SENTINEL-2
- 2A: 2015; 2B: 2016–17
- Optical sensors with 13 bands
- 2–5 days
- Monitoring land-use changes, agriculture and ecosystems, volcanoes and landslides

SENTINEL-3
- 3A: 2015, 3B; 2016–17
- Sea/land temperature radiometer, sea/land colour instrument
- 1–2 days
- Sea-surface and land-ice topography, sea and land surface temperatures and colours

SENTINEL-4
- Near end of this decade
- Ultraviolet/visible/near-infrared spectrometer
- Geostationary. Hourly coverage of Europe/North Africa
- Monitoring of air pollution, stratospheric ozone, solar radiation

SENTINEL-5
- Near end of this decade
- Ultraviolet/visible/near-infrared/ shortwave spectrometer
- 17 days
- Monitoring of air pollution, stratospheric ozone, solar radiation and climate

SENTINEL-6
- Mission still under discussion
- Radar altimeter
- Under discussion
- Measure global sea-surface height for oceanography and climate studies

---

SPACE

Earth observation enters next phase

Expectations high as first European Sentinel satellite launches.

BY DECLAN BUTLER

Europe has launched the first satellite of what is heralded as one of the most ambitious Earth-observation programmes ever. On 3 April, a Soyuz rocket dispatched into orbit the Sentinel-1A probe, the first craft of a planned constellation of six Sentinel families set to be launched by the end of the decade. Together, the satellites will offer unprecedented long-term monitoring of the planet's land, water and atmosphere.

The Sentinels will be the core of the €8.4-billion (US$11.5-billion) Copernicus programme, which is managed by the European Commission. Copernicus will also draw in data from about 30 other satellites, and from ocean buoys, weather stations and air-quality monitoring networks.

"The Sentinels and Copernicus have the potential to become the world's most comprehensive Earth-monitoring system," says Zbynek Malenovsky, who studies vegetation using remote sensing at the University of Wollongong in Australia.

Copernicus was designed by the European Union (EU) and the European Space Agency (ESA) to help the European Commission and EU member states to develop environmental policies and monitor the results. Its data will be used to create services for myriad practical applications, including ice mapping, agriculture management, climate-change forecasting and disaster response. The idea is to produce images, maps and models in near real time, much as is done with weather monitoring, but for many more variables.

Unlike most previous Earth-observation missions, the Sentinels will be replaced regularly as they age. This will help to generate long-term cross-calibrated data sets of a variety of imagery and measurements, says Cathy Clerbaux, an atmospheric scientist at the LATMOS atmosphere and astrophysics research institute in Paris. "It's not easy to connect data series such as measurements of pollutants, ozone or greenhouse gases when you have different instruments, and gaps between missions," she says.

The data will be free for anyone to access and use. But researchers will enjoy formal user status alongside public authorities, and will thus have privileged access, including dedicated help desks and support. "Scientists are now much more integrated into the user community, and not neglected as they have been in the past, when the focus was more on the operational side," says Josef Aschbacher, head of ESA's Copernicus office. "I expect scientists to be the number-one user group."

Sentinel-1A is the first of two identical satellites; 1B is set to be launched in the next 18 months. Both contain a radar system that can see in darkness and through clouds, unlike the optical instruments on many satellites. This will allow them to continuously image cloudy areas such as tropical forests. They will operate in tandem, cutting down the time between flyovers of the same point on Earth (known as revisit time), and enabling quick-succession imaging to measure, for example, ground deformation from earthquakes.

Sentinels 2 to 5 will have different goals. Between them, they will use optical sensors, radiometers and spectrometers to measure everything from sea temperatures to air pollution. In addition, a Sentinel-5 Precursor satellite will be launched in 2016 to minimize the shortfall in atmospheric data-gathering following the 2012 loss of the European Envisat satellite. A sixth Sentinel, a radar altimeter that will measure sea-surface heights, is also under discussion (see 'Watchers in the skies').

These diverse measurements of the major components of Earth systems will make the Sentinels very valuable, says Richard Anthes, emeritus president of the University Corporation for Atmospheric Research in Boulder, Colorado. "A balanced suite of Earth observations is required for observing and understanding Earth as an interconnected system," he says.

Sentinel-4, for example, will be one of the first satellites to monitor atmospheric

pollutants from a geostationary orbit, notes Clerbaux — and the first to provide hourly measurements over a single area, in this case most of Europe and North Africa.

Sentinel-2, a pair of high-resolution imaging devices, is also causing excitement. The satellites' specifications are superior to those of Landsat-8, the flagship US Earth-observation satellite, with a spatial resolution down to 10 metres — three times finer than Landsat-8 — and shorter revisit times of just 2–3 days at mid-latitudes. This opens up research into areas that

update every few days, such as crop changes.

"Sentinel-2 should really change the face of Earth observing," says Gregory Asner, an Earth scientist at the Carnegie Institution for Science in Stanford, California. "This is the satellite that could revolutionize land-cover and land-use change monitoring and analysis."

Scientists from Sentinel-2 and Landsat-8 have been working together to make their data compatible and to develop joint archives. It is a test of the concept of a virtual satellite constellation, says Mike Wulder, a scientist at

the Canadian Forest Service in Victoria and a member of the Landsat science team. "Satellite data products could be significantly improved if these were not limited to individual sensors but would combine complementary platforms across space agencies and sensor types."

Compatibility, says Malenovský, will be a key factor within the Sentinel fleet. The fleet's scientific value, he says, will be maximized if data from various crafts can be combined to create virtual, as well as practical, constellations. ■ **SEE EDITORIAL P.149**

POLICY

Funders punish open–access dodgers

Agencies withhold grant money from researchers who do not make publications openly available.

BY RICHARD VAN NOORDEN

For years, two of the world's largest research funders — the US National Institutes of Health (NIH) and the Wellcome Trust in the United Kingdom — have issued a steady stream of incentives to coax academics to abide by their open-access policies.

Now they are done with just dangling carrots. Both institutions are bringing out the sticks: cautiously and discreetly cracking down on researchers who do not make their papers publicly available.

Neither agency would name those who have been sanctioned. But the London-based Wellcome Trust says that it has withheld grant payments on 63 occasions in the past year because papers resulting from the funding were not open access. And the NIH, in Bethesda, Maryland, says that it has delayed some continuing grant awards since July 2013 because of non-compliance with open-access policies, although the agency does not know the exact numbers.

The result, say officials, has been a noticeable jump in researchers following the rules. The NIH's compliance rate — the percentage of papers placed in the PubMed Central database for public access no later than a year after publication — now stands at 82% (see 'Opening up'). It had flatlined at around 75% for two years, says Neil Thakur, who oversees policy for the NIH's Office of Extramural Research. The Wellcome Trust's compliance

rate is 69%, up from 55% in March 2012, says Robert Kiley, head of the trust's digital services.

The stricter enforcement by the Wellcome Trust began in June 2012, when its then-head Mark Walport (now the UK government's chief science adviser) said that it was "simply unacceptable" that almost half of publications resulting from the trust's funding remained behind paywalls. The trust had mandated since 2006 that the results of work it funded should be made public, yet it had never enforced the policy. In November 2012, the NIH announced a similarly tough line on its own public-access policy; that, too, had never been enforced, despite being a legal requirement since 2008.

The NIH and the Wellcome Trust are the only funders in the world to withhold grants for open-access violations so far. Funders in other nations that have open-access policies — such

as Germany, France and Australia — do not generally track their compliance rates. These other agencies may start to mimic the tougher stances of the Wellcome and the NIH, says Peter Suber, director of the Harvard Open Access Project in Cambridge, Massachusetts.
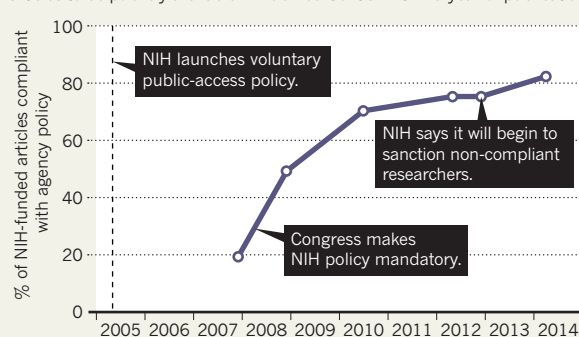
Withholding grants is not the only enforcement tool available to funders. On 31 March, four UK higher-education funders announced that from 2016, only open-access papers posted to online institutional archives will be considered in the Research Excellence Framework — the periodic research audit that grades academic work and guides the distribution of money to UK universities. Because the audit has become an academic obsession (see *Nature* **502**, 288–290; 2013), the policy will be "a game-changer", says Paul Ayris, director of library services at University College London.

Research institutions can also help with bottom-up policy enforcement, says Bernard Rentier, rector of the University of Liège in Belgium. At Liège, only articles placed in a local repository count towards internal evaluations such as pay rises and promotions. Rentier says that almost 50% of Liège's publications are now accessible. By contrast, at the Massachusetts Institute of Technology in Cambridge, which has an open-access policy it does not enforce, just 37% of papers published since 2009 are openly available from the local repository.

Without the buy-in of researchers, funding agencies know that no amount of whip-cracking will be enough, says Kiley. Some scientists are not even aware that they could be penalized. *Nature*'s news team contacted Sheila MacNeil, a tissue engineer at the University of Sheffield, UK, who has published hundreds of articles, including a March 2013 paper on making stem-cell lattices for corneal repair that was funded by the Wellcome Trust (I. Ortega *et al. Acta Biomater.* **9,** 5511–5520; 2013). *Nature* pointed out that the article should be open access but is not. "This is new to me," responds MacNeil, who plans to make the paper available. "Agreeing with open access is easy — making it happen, less so," she says. ■ **SEE GO.NATURE.COM/SSKWDW FOR A Q&A WITH ROBERT KILEY.**

OPENING UP

More researchers are abiding by the open-access policy of the US National Institutes of Health (NIH) since enforcement was stepped up. It requires articles to be publicly available in PubMed Central within a year of publication.

- NIH launches voluntary public-access policy.
- NIH says it will begin to sanction non-compliant researchers.
- Congress makes NIH policy mandatory.

% of NIH-funded articles compliant with agency policy

POLITICS

Western science severs ties with Russia

Country's science renaissance threatened as NATO and NASA suspend links.

BY QUIRIN SCHIERMEIER

Scientific relations between Russia and the West have reached their lowest ebb since the cold war, after Russia annexed Ukraine's Crimean peninsula. The North Atlantic Treaty Organization (NATO) and NASA both cut ties with Russia last week.

On 1 April, NATO suspended all civilian and military cooperation with Russia. This affects scientific collaboration under the organization's Science for Peace and Security Programme, which underpins counterterrorism and disaster-relief work, including work on technology that can detect hidden bombs at crowded public-transport locations. NATO is looking for other partners to continue these projects.

And on 2 April, the US government suspended all contact between NASA and Russian space agencies and government representatives — including visits, meetings and even e-mails. Only activities involving the International Space Station are permitted to continue. The space station has both Russian and US crew-members on board, and has relied on Russian Soyuz craft for transport since the United States retired its shuttle fleet in 2011.

It is unclear how other space-science collaborations will be affected. Igor Mitrofanov of the Space Research Institute in Moscow, who led work on a neutron-detector device on NASA's Mars Curiosity rover, says that he will not comment on the issue until NASA has notified him of the situation.

The United States and the European Union (EU) have also imposed sanctions against several high-ranking Russian government officials. The US list includes former science minister Andrei Fursenko, who acts as a personal science adviser to Russian President Vladimir Putin.

The United States, the EU and Russia have embarked on diplomatic efforts to defuse the situation in Crimea following a crisis meeting in Paris on 30 March between US secretary of state John Kerry and Russian foreign minister Sergey Lavrov. If those efforts fail, Russia's increasing isolation could become a serious long-term problem for Russian and international science, says Harley Balzer, who specializes in international affairs and Russian politics at Georgetown University in Washington DC.

"If Russia were to push one inch further into Ukraine, cutbacks on all sorts of academic exchange programmes and scientific collaborations will inevitably follow," says Balzer. Affected schemes could include the US Fulbright Program, which funds scholarly exchanges with several countries, including Russia.

Further sanctions, Balzer adds, would thwart Russia's efforts to strengthen its research and education systems and to attract foreign talent. In his 2012 election campaign, Putin promised to create several 'world-class' universities by 2020, and to substantially raise science spending — it currently stands at a mere 1.3% of gross domestic product. "Putin is killing the chance to make up lost ground," says Balzer.

The symbol of Russia's science aspirations is the Skolkovo Institute of Science and Technology (Skoltech), an English-language research university being created on the outskirts of Moscow in partnership with the Massachusetts Institute of Technology (MIT) in Cambridge.

MIT is developing curricula and research programmes and providing administrative support to Skoltech. It is also running the international calls for proposals for Skolkovo's 15 planned research centres of excellence; 6 centres already exist and 4 more are to be created this year. The terms of MIT's contract with Skoltech are confidential, but sources say that Russia is paying the US institute at least US$300 million (see *Nature* **500,** 262–264; 2013). Balzer predicts that MIT will come under "enormous pressure" to shut down the collaboration should the Crimea crisis escalate.

Edward Crawley, an MIT engineer and the president of Skoltech, says that MIT and Russian officials wish to continue the partnership, and that plans for the four new centres are moving ahead. "Conveying the idea of Skoltech takes on additional significance in these times of strained relations," says Crawley. "When the seas between two countries are stormy, it is the role of scientists and educators to put ballast to the bottom of the ship."

Attracting scientists from abroad is crucial for Russia's scientific renaissance, says Irina Dezhina, a science-policy analyst at the Institute of World Economy and International Relations in Moscow, who heads a research group at Skoltech. A series of collaborative events has been planned to run throughout 2014 for the EU–Russian Year of Science. What happens now depends on the West, she says.

People might think twice about going to a country that violates international law, warns Oleg Kharkhordin, rector of Russia's European University at St Petersburg. "But," he adds, "it should really be the interest of both sides to foster free scientific exchange." ■ **SEE EDITORIAL P.149**

NASA

The International Space Station is exempt from NASA's restrictions on formal dealings with Russia.

JESSICA BRYANT

Fewer than 30 Hainan gibbons remain today.

CONSERVATION

Time running out for rarest primate

Rescue bid launched to save Hainan gibbon from becoming first ape driven to extinction by humans.

BY DANIEL CRESSEY

China's wildlife conservation efforts are under scrutiny as scientists battle to save a species found only in a tiny corner of an island in the South China Sea. The Hainan gibbon is the world's rarest primate and its long-term survival is in jeopardy, according to an analysis.

Only 23 to 25 of the animals are thought to remain, clustered in less than 20 square kilometres of forest in China's Hainan Island. The species (*Nomascus hainanus*), which numbered more than 2,000 in the late 1950s, has been devastated through the destruction of habitat from logging, and by poaching. Extinction would give the gibbon the unwelcome distinction of being the first ape to be wiped out because of human actions. To hammer out a plan to save it, international primate researchers convened an emergency summit in Hainan last month.

"With the right conservation management, it is still possible to conserve and recover the Hainan gibbon population," says meeting co-chair Samuel Turvey, who studies animal extinctions at the Zoological Society of London (ZSL). "But given the current highly perilous state of the species, we cannot afford to wait any longer before initiating a more proactive and coordinated recovery programme." He adds that the meeting was a successful first step towards saving the animal and that a plan of action is being finalized.

The plan will be based in part on a 'population viability analysis' that models the potential size of the gibbon population in coming decades for a range of different scenarios. It is being drawn up by Kathy Traylor Holzer, a conservation planner at the Conservation Breeding Specialist Group in Apple Valley, Minnesota. "It's one of the smallest populations I've ever worked with," says Traylor Holzer. "That number — in one place — is extremely scary."

Preliminary modelling, which considers factors such as breeding success, habitat changes and natural threats, suggests that the Hainan gibbon may be safe from extinction in the next couple of decades. But its restricted habitat means that a single catastrophic event, such as a typhoon or a disease outbreak, could wipe out the minuscule population. Furthermore, low genetic diversity in the remaining animals could result in unhealthy offspring because of inbreeding. To better understand the genetics of the animals, ZSL researchers are conducting DNA sequencing using collected faeces.

Another potential problem is that the gibbons originally lived mainly in lowland forest, but logging has driven them to a higher altitude. Some scientists think that their current home might not be optimal for their needs, for example because it does not provide an adequate year-round supply of the fleshy fruits they prefer to eat.

Increased monitoring may be the key to the gibbons' conservation. Spotting the animals can be difficult, however, so researchers sometimes rely on the apes' social calls to find them. And Turvey notes that when young male gibbons leave the established social groups — a normal part of their life cycle after they reach sexual maturity — they tend to make fewer calls and are therefore difficult to find.

Bosco Chan, a specialist in Chinese conservation at the Kadoorie Farm and Botanic Garden in Hong Kong, says that one of the biggest challenges is finding enough suitable forest for the Hainan gibbon's population to expand into. Chan, who studies the gibbons and attended the meeting, notes that conservation in Hainan would also benefit other species: the gibbon is an 'umbrella species', meaning that measures taken to safeguard it will also protect the wider ecosystem of the island.

China has suffered major biodiversity losses through its rapid industrialization in recent decades. Although there are some notable conservation successes — such as for the giant panda — there have been major setbacks too; for example, the Yangtze river dolphin (*Lipotes vexillifer*) has been considered 'functionally extinct' since 2007 (see *Nature* http://doi.org/b2x2fh; 2006). Researchers hope that the Hainan gibbon will join the panda as an icon of conservation, reflecting China's renewed interest in protecting its wildlife.

"The government is paying more attention; the general public is paying more attention," says Chan. Turvey agrees, noting that representatives from a number of Chinese governmental institutions were at the workshop.

Chan is hopeful because he has seen more gibbons in recent years. "I'm quite positive. When I first started back in 2003, there were only 13 individuals we could confirm," he says. "Last year, we recorded 23." ∎

CORRECTION

The affiliation for Joshua Tewksbury in the Editorial 'Natural decline' (*Nature* **508,** 7–8; 2014) should have said he is director of the Luc Hoffmann Institute at the conservation group WWF in Gland, Switzerland.

Holding back the tide

....

With the Ganges–Brahmaputra delta sinking, the race is on to protect millions of people from future flooding.

....

BY QUIRIN SCHIERMEIER

In the wake of Cyclone Aila in 2009, swollen seas washed over the delta of the Ganges and Brahmaputra rivers. The storm surge breached the embankments surrounding a small island that was home to 10,000 families, turning the land into a muddy hell. The deluge of salty water washed out fields, homes, roads and markets just as people had begun to recover from the damage caused 18 months before by Cyclone Sidr. Many migrated to nearby cities. And thousands more took shelter on what remained of the embankments, where lack of sanitation and privacy would soon spur disease and crime.

When Steven Goodbred, an Earth and environmental researcher at Vanderbilt University in Nashville, Tennessee, came across this site during a field trip in 2011, he and his students were shocked to find the land still badly flooded and thousands of families living in tents and ramshackle huts. The broken embankments had been poorly repaired and the homesteads that they were supposed to protect remained uninhabitable. "It looked as desolate as the Moon — mud everywhere," says Goodbred. "I'd never seen anything like it." He made it his mission to determine how the embankments around the island, called Polder 32 after the Dutch word for land protected by dikes, had been eroded and undermined enough for a relatively small storm to wreak such havoc.

Scenes of disaster are not unusual in Bangladesh. About 6,000 square kilometres of the massive Ganges–Brahmaputra delta, the largest delta in the world, lies less than two metres above sea level. On average, 6,000 people in Bangladesh die each year in storms and floods. In April 1991, a single cyclone, the worst in recent decades, wiped out well over 100,000 lives in the delta and left millions of people homeless.

Risks are expected to climb. Global warming is raising sea levels around the planet by 2–3 millimetres each year. That only adds to bigger problems in the Ganges–Brahmaputra delta, which is sinking so rapidly that the local, relative sea level may be rising by up to 2 centimetres each year. And Bangladesh's population of more than 150 million people is projected to grow by a further 50 million by 2050, putting more people in harm's way.

Gloomy forecasts warn that millions of Bangladeshis might be displaced by the end of the century. Yet scientists such as Goodbred see a ray of hope. Last month, the country joined with the Netherlands to launch the Bangladesh Delta Plan, to protect the area's people from flooding. To shore up the specifics, researchers are now scrambling to provide basic data on just how fast the delta is sinking and why, and how best to guard against or even stop it.

GOING UNDER

The Ganges–Brahmaputra delta is the dumping ground of the Himalayas. As wind and rain erode the mountain range, massive rivers carry more than a billion tonnes of sediment into the Bay of Bengal each year; in some places, the layer deposited since the most recent ice age is more than one kilometre thick. As in all deltas, this loose material compacts easily, causing the land to sink slowly and the relative sea level to rise. In the past, sediment carried downstream each year would have refreshed the delta. But agriculture, industry and hydroelectric dams have diverted water and choked the flow of sediments, so the land is no longer being rebuilt. A 2009 study found[1] that

A SINKING DELTA

River sediments help to build up the Ganges–Brahmaputra delta — but not enough to counteract soil compaction and sea-level rise. Different studies show how fast the land is sinking from place to place.

SATELLITE DATA
Relative sea level rise of 8–18 mm per year.

STUDY OF ANCIENT TEMPLES
Sinking of 1–2.5 mm per year.

GPS STATIONS
Sinking of 9 mm per year in the west, and 2–4 mm per year in the east.

TIDE GAUGES
Mean sea level is climbing at 5 mm per year, but high-tide level has risen by 15–20 mm per year in some spots.

STUDY OF ANCIENT KILNS
Sinking of tens of centimetres during earthquakes.

Brahmaputra River · Ganges River · Dhaka · Chittagong · Bay of Bengal · Polder 32 · BANGLADESH · INDIA

✕ Global Positioning System (GPS) monitoring sites
● Tide gauges
▨ 1 part per thousand salt in groundwater (the official limit for US drinking water), or more

100 km

Bangladeshis attempt to rebuild a flood barrier destroyed by Cyclone Aila in 2009.

85% of the world's largest deltas suffered severe flooding in the first decade of the twenty-first century. Under current projected rates of sediment sinking and sea-level rise, the area of land at risk on deltas globally is expected to increase by at least half by 2100.

Previous efforts at flood defence in Bangladesh have not solved the problems. In 1990, the first flood action plan called for barriers to be built along main rivers: in less than 10 years, some 3,500 kilometres of embankments sprang up. In 2000, the country's focus shifted towards constructing more storm shelters and improving warning systems — but even so, about three-quarters of Bangladesh's population remains exposed to severe flooding.

A big part of the problem is a lack of understanding of how the delta's behaviour differs from one place to another. Embankments might work to some extent to protect the capital, Dhaka; but as Polder 32 demonstrates, they do not always do their job elsewhere. "Everyone who says something simple about a big delta has never been to one," says James Syvitski, a geologist at the University of Colorado Boulder. For subsidence rates in Bangladesh, he says, "depending on how and where you measure, you might get 15 different values".

Things are dramatically different between the west and east sides of the delta, for example. Over the past several centuries, geological forces and erosion have shifted the lower stretch of the Ganges steadily to the east, leaving the western parts of the delta especially starved of sediment. That makes the southwest particularly vulnerable to both seawater flooding and intrusion of salt into groundwater, which can make the water unfit to drink (see 'A sinking delta'). Researchers need to quantify and map that complexity if policy-makers are to stand a chance of addressing the problems, says Catharien Terwisscha van Scheltinga, a water-management researcher at Wageningen University in the Netherlands who is helping to prepare the Bangladesh Delta Plan.

Syvitski's work with satellite data[1] suggests that the delta is sinking below sea level by between 8 and 18 millimetres per year. But those numbers need to be checked against ground-based measurements — which are now on the way. Michael Steckler, a geologist at Columbia University's Lamont–Doherty Earth Observatory in Palisades, New York, has been installing a network of Global Positioning System (GPS) receivers to monitor subsidence since 2003. He currently maintains around 20 sites, including one established on Polder 32 last year. So far, the results suggest subsidence rates of some 9 millimetres per year in the southwest, and just 2–4 in the southeast. But the sites are still few and far between, and there are not many in the most vulnerable locations.

Some flood-control efforts might exacerbate the problems. UK geographers John Pethick at Newcastle University and Julian Orford at Queen's University Belfast reported[2] last year that water levels at some spots in the most vulnerable southwest are much higher than expected. They concluded that embankments along hundreds of tidal channels, some of which reach hundreds of kilometres inland, have vastly reduced the area of land covered by water at high tide. Because the water is less able to spread out, it shoots farther inland where it can. The result is a huge increase in tidal range in less-protected areas. Tide-gauge records

from three locations in the southwest suggest a mean rate of relative sea-level rise of about 5 millimetres per year over the past 30 years, but some local spots have experienced an average annual high-water-level increase of 15–20 millimetres.

That tidal amplification has big implications for coastal protection, says Goodbred: for areas with long tidal channels, building more embankments could actually cause higher tides and exacerbate salt-water intrusion. But others warn that it is hard to say whether Pethick and Orford's findings will hold true across the southwestern coast — it might depend strongly on the steepness of the channel walls, for example. Extrapolating is "misleading", says Maminul Haque Sarker, a geologist and deputy executive director of the Center for Environmental and Geographic Information Services in Dhaka.

Sarker points to parts of the delta that seem to be sinking much slower. In 2012, he and his colleagues took measurements to see how far the bases of ancient mosques and temples have been buried beneath incoming sediments. The present plinth levels suggest a subsidence rate of just 1–2.5 millimetres per year. Sea-level rise would have to be added to that to get a relative sea-level change. But Sarker thinks that the entire delta is sinking less significantly than researchers such as Syvitski have proposed.

EARTH-SHAKING CHANGE

Whatever the current rate of subsidence, it may not reflect the full scale of the problem. In 2010, Till Hanebuth, a geologist at the University of Bremen in Germany, excavated[3] more than a dozen ancient kilns in the Sundarbans, a coastal region of mangrove forests renowned for its population of royal Bengal tigers. The kilns were built for salt production some 300 years ago, just above the winter high-tide level of the time. Today they are buried 1.5 metres beneath the mud and the modern sea level, indicating an average sinking rate of about 5 millimetres per year.

Hanebuth thinks that the drop happened not at a slow, constant pace, but in a succession of abrupt events related to big earthquakes or cyclones. Mud-filled stumps in the area show that mangrove trees died from flooding around 1676 and 1762, when strong earthquakes hit the region. The quake in 1762, estimated at magnitude 8.8, caused land around the southeastern city of Chittagong to sink by several metres; in the Sundarbans it seems to have caused at least a 20-centimetre drop, says Hanebuth. Seismologists think that another major quake is overdue in the tectonically unstable region, and that when it comes it will devastate poorly built high-density cities such as Dhaka and Chittagong. It could also cause patches of the delta to drop more in one fell swoop than they have over decades of slow sea-level rise and sediment compaction.

Other complicating factors are easier to assess and guard against. A year after their initial observations, Goodbred and his team returned to Polder 32 armed with GPS receivers and survey tools. They found that the embankments that protect the land from the river and sea had also robbed it of fresh supplies of sediment: during the five decades of its existence, the polder had sunk by a full metre relative to the land outside the embankments because it was not being replenished. On top of that, Goodbred found, local shrimp farmers had drilled holes into the dikes to pipe salt water from coastal rivers into their hatcheries, weakening the barriers.

Although Cyclone Aila brought much suffering to the people, it helped to rescue the land a little. During the two years in which the dikes were broken, the polder rebounded with tens of centimetres of sediment deposited by daily tides. The mud caused havoc in the short term by flooding peoples' floors and gardens, but offers the possibility of long-term sustainability for delta ground.

All this information will feed into the Delta Plan, which will be written over the next 2.5 years by a Dutch–Bangladeshi consortium of government departments, research organizations and engineering consultants. The Netherlands has pledged an initial €7 million (US$9.7 million) to develop the strategy. "Having the Netherlands and their invaluable treasure of experience on board is a big push for our flood-defence efforts," says Shamsul Alam, head of the general economics division of Bangladesh's planning ministry, which coordinates the Delta Plan.

The uncertainty in the science makes it difficult for policymakers to see how much investment is justified, and what kind, says van Scheltinga. But at least some of the problem is coming into better focus. "We're only beginning to understand how the delta works — but we know enough to do a bit better," says Goodbred.

> " *We're only beginning to understand how the delta works — but we know enough to do a bit better.* "

In rural coastal areas, Goodbred adds, one solution might be to return to the kind of low, flexible embankments that people in this region built before the 1960s. Locals could raise them in the dry season to keep salty water away, and cut them down in the wet season to allow sediment in. Hugh Brammer, a UK geographer who consulted on Bangladesh's 1990 flood action plan for the World Bank, agrees that flexible barriers are needed. Tidal water must, from time to time, be allowed to flush embanked land, he says, to deposit sediment and thus prevent the polders from sinking over the long term. Homes in these polders, he notes, tend to be on land that sits half a metre or more above the polder basins, so would be protected from the influx.

The most urgent step, says Brammer, is to divert water from the Ganges to the western parts of the delta, so that the people there have access to fresh water in the dry season. In 2008, the Bangladesh government promised to consider one diversion scheme. But the costs and feasibility of such a major engineering project have yet to be properly examined; Alam says that such a diversion is unlikely to happen any time soon.

There are cheaper options, he says. Scientists with the Bangladesh Rice Research Institute in Gazipur have developed salt-tolerant rice to grow in flood-prone plains. In large coastal cities, new homes and public infrastructure could be built on artificially raised land. Near the coast, conserving and planting trees could create a buffer against storm surges.

In its 2011–15 economic plan, Bangladesh earmarked more than 120 billion taka (US$1.5 billion) — 4% of all public expenditure — for climate adaptation and disaster management. Furthermore, it is currently channelling a $170-million multi-donor climate-change-resilience fund, set up in 2010, into projects including flood protection. More money could come from the multi-billion-dollar international Green Climate Fund. To enact a fully fledged Delta Plan, says Saleemul Huq, a senior fellow in the climate-change group at the International Institute for Environment and Development in London, will require several billion dollars over the next few years. Bangladesh currently gets about $2 billion per year from donors for everything from economic development to food relief, so Alam says that it is not unreasonable to hope that it will be able to redirect funds to achieve the flood goals.

Meanwhile, life on Polder 32 is slowly returning to normal. Not all of the destroyed embankments have been repaired or replaced, but hundreds of homesteads and businesses are being moved to higher land both inside and outside the polder. Goodbred trusts that the delta and its people have a future. "Bangladesh is blessed by options," he says. "But time is short and the issues are substantial." ■

Quirin Schiermeier is a senior reporter for Nature in Munich, Germany.

1. Syvitski, J. et al. Nature Geosci. **2,** 681–686 (2009).
2. Pethick, J. & Orford, J. D. Glob. Planet. Change **111,** 237–245 (2013).
3. Hanebuth, T. J. J., Kudrass, H. R., Linstädter, J., Islam, B. & Zander, A. M. Geology **41,** 987–990 (2013).

The clock–watcher

*Biomathematician **Steve Horvath** has discovered a strikingly accurate way to measure human ageing through epigenetic signatures.*

BY W. WAYT GIBBS

As a teenager in Germany, Steve Horvath, his identical twin Markus and their friend Jörg Zimmermann formed 'the Gilgamesh project', which involved regular meetings where the three discussed mathematics, physics and philosophy. The inspiration for the name, Horvath says, was the ancient Sumerian epic in which a king of Uruk searches for a plant that can restore youth. Fittingly, talk at the meetings often turned to ideas for how science might extend lifespan.

At their final meeting in 1989, the trio made a solemn pact: to dedicate their careers to pursuing science that could prolong healthy human life. Jörg set his eye on computer science and artificial intelligence, Markus on biochemistry and genetics, and Steve says that he "planned to use mathematical modelling and gene networks to understand how to extend life". Jörg did end up working in artificial intelligence, as a computer scientist at the University of Bonn in Germany, but "Markus fell off the wagon", his brother says, "and became a psychiatrist".

Steve, now a human geneticist and biostatistician at the University of California, Los Angeles (UCLA), says that he finally feels poised to make good on the promise. Through a hard-fought project that involved years of solo work, multiple rejections by editors and reviewers and battling through the loss of a child, he has gathered and analysed data on more than 13,000 human tissue samples[1]. The result is a cellular biological clock that has impressed researchers with its accuracy, how easy it is to read and the fact that it ticks at the same rate in many parts of the body — with some intriguing exceptions that might provide clues to the nature of ageing and its maladies.

Horvath's clock emerges from epigenetics, the study of chemical and structural modifications made to the genome that do not alter the DNA sequence but that are passed along as cells divide and can influence how genes are expressed. As cells age, the pattern of epigenetic alterations shifts, and some of the changes seem to mark time. To determine a person's age, Horvath explores data for hundreds of far-flung positions on DNA from a sample of cells and notes how often those positions are methylated — that is, have a methyl group attached.

He has discovered an algorithm, based on the methylation status of a set of these genomic positions, that provides a remarkably accurate age estimate — not of the cells, but of the person the cells inhabit. White blood cells, for example, which may be just a few days or weeks old, will carry the signature of the 50-year-old donor they came from, plus or minus a few years. The same is true for DNA extracted from a cheek swab, the brain, the colon and numerous other organs. This sets the method apart from tests that rely on biomarkers of age that work in only one or two tissues, including the gold-standard dating procedure, aspartic acid racemization, which analyses proteins that are locked away for a lifetime in tooth or bone.

"I wanted to develop a method that would work in many or most tissues. It was a very risky project," Horvath says. But now the gamble seems to be paying off. By the time his findings were finally published last year[1], the clock's median error was 3.6 years, meaning that it could guess the age of half the donors to within 43 months for a broad selection of tissues. That accuracy improves to 2.7 years for saliva alone, 1.9 years for certain types of white blood cell and 1.5 years for the brain cortex. The clock shows stem cells removed from embryos to be extremely young and the brains of centenarians to be about 100.

"Such tight correlations suggest there is something seemingly immutable going on in cells," says Elizabeth Blackburn of the University of California, San Francisco, who won a Nobel prize for her research on telomeres — caps on the ends of chromosomes that shorten with age. It could be a clue to undiscovered biology, she suggests. And there may be medical implications in cases in which epigenetic estimates do not match a person's birth certificate.

In the months since Horvath's paper appeared, other researchers have replicated and extended the results. The study has stirred up excitement about potential applications, but also debate about the underlying biology at work.

"It's something new," says Peter Visscher, chair of quantitative genetics at the University of Queensland in Australia. "If he's right that there is something like an inherently epigenetic clock at work in ageing, that is very interesting. It must be important."

CLOCKING ON

Horvath kept his vow to the Gilgamesh project by supplementing his PhD in mathematics with a doctorate in biostatistics, which led to a position in the genetics department at UCLA in 2000. After receiving tenure in 2006, he began to focus on ageing by searching for shifts that occur in gene activity over the course of life. A doctoral student took the lead, feeding gene-transcription data through statistical filters in the hope of turning up a robust biomarker for age. But after more than a year, Horvath and the student had found no strong clues. If any such signal exists in gene-transcript data, they concluded, it is hopelessly swamped by the noisy variations from organ to organ and person to person. With little to show for their work, "I decided to keep quixotic projects like this away from students and postdocs," Horvath says. "It didn't seem fair to risk their careers."

Things began to look up in 2011, however. As part of a team led by his UCLA colleague Eric Vilain, Horvath had analysed methylation patterns in DNA extracted from the saliva of 68 adults. The researchers were looking for an epigenetic pattern that correlated with sexual orientation. None turned up, but with the data in hand Horvath and his colleagues decided to see whether they could use it to predict age.

In human DNA, methyl groups most often attach at 'CpG sites' — places where a cytosine precedes a guanine in the DNA. A typical human genome contains more than 28 million such sites. But the microarray technology used to detect methylation samples finds only a fraction of them: older machines pin down just 27,000 sites and newer ones around 485,000.

Horvath got lucky. He found success with a simple statistical model, which looked at how many cells in a drop of saliva have DNA methylated at just two particular CpG sites. The index roughly paralleled participants' ages with a correlation of 0.85, or 85%, and an average accuracy of about five years[2].

While working on a subsequent study, Horvath identified methylation patterns that hewed even more closely to age in very different cell types, such as brain and blood. Suddenly, a goal that he had thought impossible — finding a biomarker for the age of almost every part of the body — seemed attainable.

But it would not be easy. He would have to pull together myriad data sets that included both peoples' ages and their DNA methylation information. Methylation profiles are used for many kinds of medical research — usually in areas other than ageing (see *Nature* **508,** 22; 2014). And because of variations in the way they are collected and processed, they can be tricky to compare. Horvath worried: "How do you make data sets comparable if they were generated by different labs using different protocols?"

Building on work by Andrew Teschendorff at University College London, Horvath devised a way to normalize methylation profiles and put them all on the same footing. Beyond that, his audacious strategy for dealing with some of the uncertainty was to ignore it — and hope that it didn't clobber the accuracy of his model.

It didn't. By early 2012, his algorithm was using 16 CpG sites in the genome, and was returning correlations with chronological age of 96% in nine kinds of tissue. The accuracy was astonishing: median errors were within three years for blood samples and just 18 months for cheek swabs.

But the editors of two journals rejected Horvath's paper. The "tenor of the reviewers was that it was just too good to be true," he says. They suspected that the clock model fit the training data used to build it but that Horvath had insufficient test data to validate it thoroughly.

Humbled but undaunted, Horvath continued collecting data sets and expanding the algorithm. By December 2012, his methylation database spanned 51 types of non-cancerous tissue and cells, plus 20 kinds of cancer. The age estimator had grown to include 353 CpG sites.

He had completed his analyses and was preparing to rewrite his paper from scratch when

his pregnant wife's waters broke — more than three months ahead of her due date. For the next 20 days, he barely left the chair beside her hospital bed while she and the medical staff tried to stave off infection and premature delivery.

The stress focused him. "I wrote every hour as if it was the last hour I had for finishing the article," he says. He made good progress, as did his wife and their baby. Towards the end of the third week, with Christmas approaching, "I started to feel really hopeful," he says. But suddenly the baby's heart rate shot up. After an emergency Caesarean section, the baby struggled to breathe. "The doctors made a heroic effort," Horvath says, "but she died in my hands on the day of her delivery. It wasn't until 10 days later that I found enough strength to upload the paper to *Genome Biology*."

The reviews came back in the spring: more disbelief, and another rejection. Horvath didn't blame the reviewers for being sceptical. "Everyone who develops biomarkers knows what to expect: a very strong biomarker gives you a correlation of, say, 0.6 or 0.7." For example, the correlation between age and the length of telomeres is less than 0.5. For Horvath's clock algorithm, that figure is 0.96. He confesses that he had trouble believing it himself until other researchers independently confirmed the tight association.

This time, Horvath refused to take no for an answer. "After reading the reviewers' comments, I spent the next 10 minutes doing three things that one should never do," he says. "First, I went to the fridge and drank three bottles of beer as fast as I could. Second, I went back to the computer and drafted a letter to the editor. Third, I sent it off."

ABOUT TIME

The appeal worked, and after his article[1] was featured in the October 2013 issue of *Genome Biology*, others began downloading the epigenetic-clock program from Horvath's website to test it on their own data. Marco Boks at the University Medical Centre Utrecht in the Netherlands applied it to blood samples collected from 96 Dutch veterans of the war in Afghanistan aged between 18 and 53. The correlation between predicted and actual ages was 99.7%, with a median error measured in months.

At Zymo Research, a biotechnology company in Irvine, California, Wei Guo and Kevin Bryant wondered whether the program would work on a set of urine samples Zymo had collected from 11 men and women aged between 28 and 72. The correlation was 98%, with a standard error of just 2.7 years. "That's amazingly good," Bryant says. "Urine samples weren't even part of the data that Steve used to develop this algorithm."

Horvath's method has many potential applications. Criminal investigators, for example, might find an epigenetic clock handy for establishing the age of a victim or an assailant by analysing any biological residues left behind. Trey Ideker, chief of the medical-genetics division

at the University of California, San Diego, says that his group is working with a forensics lab to test an epigenetic clock that he and his collaborators designed to work specifically on blood, using mathematical methods very similar to Horvath's[3]. Although Ideker's clock is tissue-specific and not quite as accurate, it could be cheaper to use because it is based on fewer CpG sites — 71 rather than 353.

"*The big question is whether the clock measures a biochemical process that serves a purpose.*"
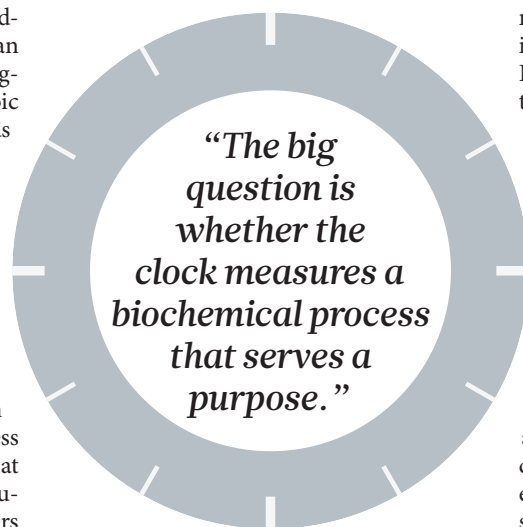
Both Ideker and Horvath expect that the most interesting use of the clock will be to detect 'age acceleration': discrepancies between a person's epigenetic and chronological ages, either overall or in one particular part of their body.

Such discrepancies could be signs that something is awry. In work due to be presented at the November meeting of the Gerontological Society of America, Brian Chen of the US National Heart, Lung, and Blood Institute (NHLBI) in Framingham, Massachusetts, teamed up with Horvath and others to analyse methylation data collected on more than 2,100 men and women aged 40 to 92 as part of the Framingham Heart Study. The researchers concluded that for every five-year increase in age acceleration, the risk of dying from any cause during the study jumped by 15%. Horvath says that unpublished work from two other large studies also finds epigenetic age acceleration to be a substantial risk factor for mortality, even after controlling for chronological age and other well-known risk factors.

Researchers are also comparing the ages of different tissues from the same individual, in the hope of identifying more accurate, less invasive ways to diagnose disease or gauge the risk of future illness. Last year, Ideker and his collaborators reported that the epigenetic ages of breast, kidney, lung and skin cancers were 40% older, on average, than the patients from which they were removed[3]. The picture from Horvath's method is less clear. Some cancers, such as brain tumours, seemed to be decades older, in terms of their methylation, than they should be. But the effect was reversed for

some other cancers, such as certain types of endometrial and breast tumours.

Distortions in epigenetic age seem to parallel other diseases more closely. Horvath says that recent work has found that people with HIV who have detectable viral loads appear older, epigenetically, than healthy people or those with HIV who have suppressed the virus. Another study, not yet published, observes that some tissues show significant age acceleration in morbidly obese people, he reports. In the coming months, he will be mining the vast Women's Health Initiative database — which includes thousands of methylation profiles gathered as part of this 20-year, 160,000-person study spearheaded by the NHLBI — for more links.

AN AGE-OLD QUESTION

Medical researchers might be able to use the epigenetic clock to better diagnose and classify illnesses even without really understanding how the biology works. But Horvath hopes that the science won't stop there.

"The big question is whether the clock measures a biochemical process that serves a purpose," he says. His best guess is that the clock corresponds to the function of an epigenomic housekeeping system, which helps to stabilize the genome by maintaining methylation patterns. The more active this mechanism, he proposes, the faster the epigenetic clock ticks.

Because methylation is usually reversible, Wei says, it might be possible to grab the minute hand of the epigenetic clock and retard its incessant progress — an idea that makes Horvath's solemn adolescent vow sound almost attainable. "The greatest hope is that this clock measures the output of a process that really does relate to ageing — even causes ageing," Horvath says.

But some are sceptical. Teschendorff's research has shown that genome-wide patterns of methylation drift gradually as the years slip past[4]. He suspects that some passive process is behind the shift and that it leads to ageing and disease mainly by interfering with the ability of stem cells to differentiate. Ideker agrees that the epigenetic transition from young to old could be mostly random, in which case there may be nothing especially informative about the 353 cogs of Horvath's clock.

Horvath acknowledges that it will take more work to find out whether epigenetic age predicts the onset of disease and decrepitude better than a calendar does. "But the epigenetic clock gives us a new start and a new hope of something that will affect ageing," he says. In that way, Gilgamesh's ancient quest for a way to delay the inevitable lives on. ●

W. Wayt Gibbs *is a freelance science writer in Seattle, Washington.*

1. Horvath, S. *Genome Biol.* **14,** R115 (2013).
2. Bocklandt, S. *et al. PLoS ONE* **6,** e14821 (2011).
3. Hannum, G. *et al. Mol. Cell* **49,** 359–367 (2013).
4. Teschendorff, A. E., West, J. & Beck, S. *Hum. Mol. Genet.* **22,** R7–R15 (2013).

COMMENT

ILLUSTRATION BY DAVID PARKINS
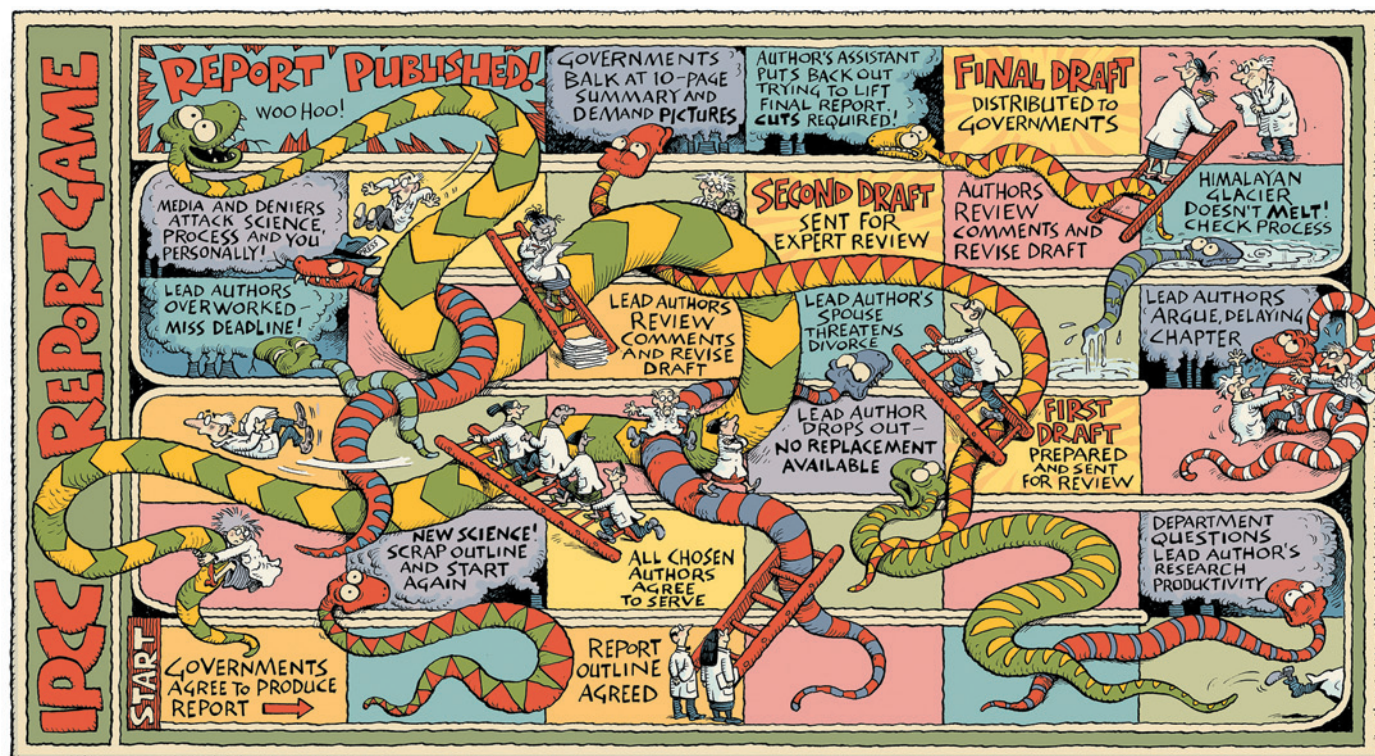
Streamline IPCC reports

As the Intergovernmental Panel on Climate Change asks how its assessment process should evolve, **Dave Griggs** argues for decadal updates and eased workloads.

Although the Intergovernmental Panel on Climate Change (IPCC) has gained a justified reputation for producing the most up-to-date, comprehensive and authoritative statements of our knowledge of climate change, this has come at a cost to the scientific community. On 13 April, the IPCC releases the last part of its Fifth Assessment Report. Each report has become longer — the latest round-up of scientific evidence is nearly four times the size of the first — and is taking more time and effort to prepare, placing a greater imposition on those involved.

Looking ahead to its sixth assessment, the IPCC has formed a task group to improve its operation and products. I believe that the panel should streamline the process and ease the pressure on the many hundreds of scientists who write, review and produce the assessments.

IPCC work can ruin lives, as a former lead author told me when I was head of an IPCC Technical Support Unit (TSU) working on the Third Assessment Report[1]. Over three years, he had devoted months of his own time to his chapter, because his university would not reduce his workload. He had haggled over details with other authors, responded to hundreds of reviewers' comments (twice) and defended the account against distortion by governments. When the report came out he was attacked by deniers and the media, causing him distress.

His marriage nearly ended. But when I asked him if he would work on the next report, he said: "Of course I will do it again, it is the most important thing I have ever done."

The IPCC reports have value for policy-making and should be continued. Although there is much that we still do not understand about the climate system, as the research has progressed, each report expresses increasing confidence in the basic science. The risks of allowing greenhouse-gas emissions to rise are clear.

But the IPCC process must be improved. Future assessments need to be shorter, the writing process streamlined, and lead authors should be compensated for their efforts. Here are seven lessons learned, often ▶

▶ the hard way, based on my involvement in the third, fourth[2] and fifth[3] assessments, and on discussions with colleagues.

SEVEN RECOMMENDATIONS

Publish shorter, less-frequent reports. With each update, IPCC reports have stretched — from 410 pages for the first Working Group I (WGI) report to 881 pages by the third and 1,535 by the fifth. There is more research to assess, but new science builds on old, so it is not necessary to go back to the beginning to build a comprehensive account. Limiting reports to 1,000 pages or fewer would save time, reduce workloads and make the reports more readable and focused.

The IPCC has produced its comprehensive reports every six years or so. From a scientific perspective, you could get the same benefit at lower cost and effort by updating them every ten years. Emerging and fast-moving areas of science could be covered in the interim using short, targeted reports.

Implementing these two changes would also enable the IPCC to maintain its small, efficient administrative structure. The panel's reports are prepared by hundreds of scientists in three working groups (WGI reports on the physical-science basis, WGII on the impacts, adaptation and vulnerability, and WGIII on the mitigation of climate change). The process is supported by a small secretariat in Geneva, Switzerland, and four TSUs (one for each working group and one Task Force on National Greenhouse Gas Inventories). Each TSU consists of just five to seven people, and it should stay that way.

Consult widely at the start. The outline for the Fifth Assessment Report was developed at a scoping meeting in July 2009 of scientists from across all three working groups, rather than by the co-chairs and the TSU, as for previous reports. This approach should be continued and extended, for example by consulting lead authors of past reports. By letting everyone have a say early in the process, we can improve consistency between the three reports, for example in how to treat uncertainties, and identify important new areas of focus, such as the contribution of ice sheets to sea-level rise. Although it is impossible to produce an outline that everyone agrees on, it is vital that everyone has an opportunity to express their opinion and understands the logic behind the report structure.

Recognize the skills lead authors need. The selection of lead authors is probably the most crucial step in the process. Lead authors oversee the drafting of the chapters and their review, first by experts, then by experts and governments. It is a monumental task. Each report can draw more than 50,000 comments, from the insightful to the stupid. Lead authors must consider each one, amend their draft to take into account valid comments and document responses to all those received.

The IPCC requires that selected authors reflect the "range of scientific, technical and socio-economic views and expertise; geographical representation; … previous experience in the IPCC; [and] gender balance". This mixture is difficult to achieve. If one lead author pulls out, it can be hard to find a replacement and it might affect the selection of other authors to redress the balance.

But lead authors need other subtle skills, some of which are rare in science. They need to be diligent, careful, balanced and fair, and able to take criticism while remaining open to changing their views in the face of new evidence. They must be amenable, dedicated enough to tolerate the tortuous process and be able to deliver on time.

"Even the most trivial sentences can cause endless debate."

Lead authors need to work to absolute deadlines, which are dictated by governments. The IPCC does not pay authors, who often work on reports in their own time, so a TSU has no means to force an author to deliver. This underlies the value of retaining some lead authors who have had previous experience in the process, and who can impress on their colleagues the importance of meeting deadlines.

Because the lead-author teams work together for years, they have to get along. When I was helping to select lead authors for the Third Assessment Report, I invited two prominent scientists to take part. Both agreed to participate — on the condition that the other person was not a lead author. Only by offering to act as a neutral arbiter in any disputes was I able to get both on board. But it is best to avoid such a situation.

Summarize with better graphics. Report authors need to develop a short summary for policy-makers and a longer technical summary. It is extremely challenging to summarize often more than 1,000 pages of dense science into just 10 or 20, let alone getting all the scientists, reviewers and governments to agree on the wording. But, for busy readers, even ten pages is too long. So a handful of key findings need to be pulled out. It is worth spending time to get these statements right, because for many people they form the primary focus. The summary for policy-makers must contain information that is relevant to them while staying faithful to the science, and should never be policy prescriptive.

A picture is worth a thousand words. Future reports would benefit from improved infographics. The reports are communication documents as much as they are scientific ones, so lead author teams should have access to professional graphics support to present their work most effectively.

Never compromise on the science. The final step in the IPCC process is the Plenary session, or meeting of governments, at which the summary for policy-makers is agreed on word-by-word. The governments can request changes to make the summary clearer or to include things that are important for policy. They cannot change the science. Lead authors are present to ensure that the summary accurately reflects the research. Such situations can be intimidating, but in my experience the science has never been compromised.

The scientists should listen, too, because government input can improve the summaries. For example, in one meeting I attended, the policy-makers rightly asked what the scientists meant by a "significant fraction" of something. Did it mean 1%, 10% or 50%, and by what criteria was it significant? We had to form a small subgroup to resolve the issue.

And even the most trivial sentences can cause endless debate. In the WGI contribution to the Third Assessment Report, we ran into a problem on the second sentence, which read: "Many hundreds of scientists from many countries participated in its preparation and review." For several hours, delegates debated what was meant by "many hundreds" and "many countries". In the end, the problem was resolved only by adding the numbers in a footnote.

I learned an important lesson in that session — patience is crucial. As the hours went by, I grew desperate, knowing that we had difficult points of science left to discuss. I asked the group co-chair, John Houghton, to try to conclude the discussion. He explained that if he cut it short he would lose the trust of the governments that still wanted to speak. If he let them all speak now, he said, they would give him more leniency later when he had to take a hard line. And so it proved.

Resist more complexity. The IPCC has strict rules and procedures, but with something so large and complex, mistakes are inevitable — although surprisingly rare. This occurred most notably in 2007 when errors were identified in the WGII contribution to the Fourth Assessment Report[4], which stated that there was a high likelihood that the Himalayan glaciers might disappear by 2035.

The question was not of whether there was error — there was — but whether this error was symptomatic of a flawed process, undermining the conclusions of all the reports. In my view, such an error does not nullify other conclusions, owing to the multiple lines of evidence presented, and I have never been involved in such a careful and rigorous process. So, although some may not like the revelations in the IPCC reports, it is wrong to shoot the messenger.

Following this incident, the InterAcademy Council reviewed IPCC procedures and

made recommendations, including more checks and balances, many of which the IPCC adopted. However, adding complexity adds to the workloads of already over-stretched scientists and will never eliminate errors completely. I believe that the process is already rigorous enough and that adding further complexity should be resisted.

Compensate lead authors for their time. Although practices vary between countries and institutions, there is no system of compensation for IPCC lead authors, other than for travel expenses. The authors devote significant amounts of time, often with no reduction in workload from their main employers. This limits the time that they can spend on the reports.

Lead authors should not profit from being involved in the IPCC, because this could be perceived as a conflict of interest. But governments should offer to cover the costs to their host institutions of teaching relief, or of employing a research assistant to ease their workload, allowing them to maintain scientific output. Although this would increase the cost of producing the reports, the extra expense would be trivial compared with the benefit of having the best possible assessment of climate change.

So why should scientists take part in IPCC assessment reports? Because the future of the world will be severely affected by climate change and we have a responsibility to make the risks known. The IPCC remains the most effective way to do that. ■

David Griggs *is professor and director of the Monash Sustainability Institute, Monash University, Melbourne, Australia. He was head of the Intergovernmental Panel on Climate Change Working Group I Technical Support Unit for the Third Assessment Report, a review editor for the Working Group I Fourth Assessment Report and a reviewer for the Fifth Assessment Report. e-mail: dave.griggs@monash.edu*

1. Intergovernmental Panel on Climate Change. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (eds Houghton, J. T. *et al.*) (Cambridge Univ. Press, 2001).
2. Intergovernmental Panel on Climate Change. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. D. *et al.*) (Cambridge Univ. Press, 2007).
3. Intergovernmental Panel on Climate Change. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) (Cambridge Univ. Press, 2013).
4. Intergovernmental Panel on Climate Change. *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Parry, M. L. *et al.*) (Cambridge Univ. Press, 2007).

Floods brought parts of Britain to a standstill earlier this year.

Improve economic models of climate change

Costs of carbon emissions are being underestimated, but current estimates are still valuable for setting mitigation policy, say **Richard L. Revesz** and colleagues.

On 31 March, the Intergovernmental Panel on Climate Change (IPCC) released its latest report on the impacts of climate change on humans and ecosystems (see go.nature.com/ad5v1b). These are real risks that need to be accounted for in planning for adaptation and mitigation. Pricing the risks with integrated models of physics and economics lets their costs be compared to those of limiting climate change or investing in greater resilience.

Last year, an interagency working group for the US government used three leading economic models to estimate that a tonne of carbon dioxide emitted now will cause future harms worth US$37 in today's dollars[1]. This 'social cost of carbon' represents the money saved from avoided damage, owing to policies that reduce emissions of carbon dioxide.

Governments, agencies and companies use such estimates to guide decisions about how much to invest in reducing ▶

emissions. In the United States, a previous estimate[2] made in 2010 informed the stricter fuel-economy requirements for new cars. The latest value is motivating President Barack Obama's plan to impose greenhouse-gas limits on coal-fired power plants by next year. Canada, Mexico, the United Kingdom, France, Germany and Norway have used similar numbers to guide regulatory decisions, as has the International Monetary Fund to analyse fossil-fuel subsidies.

Yet the social-cost benchmark is under fire. Industry groups, politicians — including leaders of the energy and commerce committee of the US House of Representatives — and some academics say that uncertainties render the estimate useless.

As legal, climate-science and economics experts, we believe that the current estimate for the social cost of carbon is useful for policy-making, notwithstanding the significant uncertainties. The leading economic models all point in the same direction: that climate change causes substantial economic harm, justifying immediate action to reduce emissions. In fact, because the models omit some major risks associated with climate change, such as social unrest and disruptions to economic growth, they are probably understating future harms. The alternative — assigning no value to reductions in carbon dioxide emissions — would lead to regulation of greenhouse gases that is even more lax.

Instead, climate-economic models need to be extended to include a wider range of social and economic impacts. Gaps need to be filled, such as the economic responses of developing countries and estimates of damages at extreme temperatures. Today, only a handful of researchers in the United States and Europe specialize in such modelling. A broader programme involving more people exploring more phenomena is needed to better estimate the social cost of carbon and to guide policy-makers. Otherwise policies will become untethered from economic realities.

SOCIAL COST

The models in question aim to integrate estimates of the costs of greenhouse-gas emissions and of steps to reduce them. First, they translate scenarios of economic and population growth, and resulting emissions, into changes in atmospheric composition and global mean temperature. Then the models apply 'damage functions' that approximate the global relationships between temperature changes and the economic costs from impacts such as changes in sea level, cyclone frequency, agricultural productivity and ecosystem function. Finally, the models translate future damages into present monetary value.

"The models omit social unrest and disruptions to economic growth."

Sources of uncertainty are numerous[3]. They include: how the climate responds to carbon dioxide concentrations; positive and negative feedback loops in the climate system; emissions growth rates for various socio-economic scenarios; the completeness and accuracy of damage functions (especially with regard to catastrophic harms, migration and conflict, weather variability and feedbacks on economic growth); the ability of future generations to adapt to climate change; and the economic 'discount rate' used to translate future costs to current dollars.

The 2013 US analysis[1] used the then-most recent vintages of three long-standing models: FUND 3.8, DICE 2010, and PAGE09. Each model applies different climatic and economic functions to simplify the complex picture. Despite the range of approaches and uncertainties, each one predicted sizeable economic damage from greenhouse-gas emissions for warming beyond 2 °C above pre-industrial levels. Two models, ENVISAGE and CRED, published since the US analysis was structured in 2010, have broadly similar projections to these three (see 'Carbon's costly legacy'). The analysis suggested that — depending on assumptions about how future damages are valued in today's money — the expected global cost of one tonne of carbon dioxide emitted in 2020 is between $12 and $64 (with $43 as the central value).

GREATER HARM

The future costs of climate change could be even higher, for four reasons. First, the impacts of historic temperature changes suggest that societies and economies may be more vulnerable than current models predict and that weather variability is more important than average weather in determining impacts, particularly for crop growth and food security. For example, the yields of some crops may decline rapidly above certain temperatures[4].

Second, the models omit damages to labour productivity, to productivity growth, and to the value of the capital stock, including buildings and infrastructure. By lowering the annual growth rate, these damages

---

CARBON'S COSTLY LEGACY

Economic models of climate change project that resulting damage worldwide (**A**) will increase with future emissions and may cost several per cent of global gross domestic product (GDP) with the warming expected by 2100. Uncertainties in future socio-economics, emission rates and climate impacts result in a range of estimates of the social cost of carbon, which is also affected by the choice of 'discount rate' used to convert future harms into today's money (**B**).

A PROJECTED DAMAGES

Damages (% global GDP) vs Global mean temperature rise (°C above pre-industrial levels)

Model
- DICE 2010
- FUND 3.8*
- PAGE09*
- CRED 1.4*
- ENVISAGE

Models predict rising damages from warming.

B SOCIAL COSTS FROM US GOVERNMENT ANALYSIS
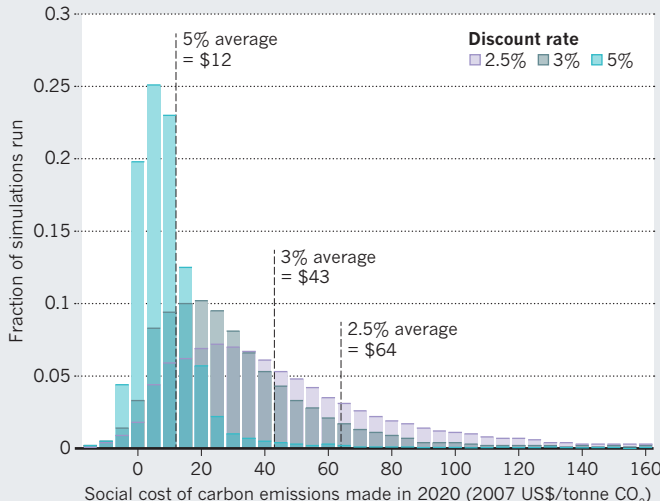
Fraction of simulations run vs Social cost of carbon emissions made in 2020 (2007 US$/tonne CO$_2$)

Discount rate: 2.5% 3% 5%

5% average = $12
3% average = $43
2.5% average = $64

*Shaded regions indicate 5% and 95% confidence intervals for FUND 3.8 and PAGE09, and a high–low range for CRED 1.4.

Storms caused chaos on roads in northwestern Italy in 2011.

could have deeper and longer-lasting effects on the global economy than the static losses of annual economic output currently represented in the three main models[5,6]. A significant decline in human welfare is possible in the medium and long run owing to the compounding effects of lost growth. Also not taken into account are the risks of climate-induced wars, coups or societal collapses and the resulting economic crises[7].

Third, the models assume that the value that people attach to ecosystems will remain constant[8]. Yet as a commodity becomes more scarce, its value increases. In the desert, water is extremely valuable. During a flood, dry land is highly prized. Because the services provided by ecosystems are likely to decline as warming degrades them, the costs of future ecosystem damage from climate change will rise faster than the models predict.

Fourth, the US analysis assumes a constant discount rate to translate future harms into today's money. However, for impacts that are both highly uncertain and occurring in the distant future, economists have shown[9] that a discount rate that declines over time should be used, with discount rates for the far future significantly below those that were used in the 2013 analysis. This approach would

yield a higher present value to the long-term impacts of climate change and thus a higher value for the social cost of carbon.

It is true that future technological developments might better equip society to cope with climate change. And of course overall bias cannot be determined simply by adding biases in each direction. But the bulk of the literature and arguments indicates that social-cost models are underestimating climate-change harms.

BETTER MODELS
What now? Modellers, scientists and environmental economists must continue to step outside their silos and work together to identify research gaps and modelling limitations.

Climate hot spots in the developing world are one such gap, because economic responses in these regions cannot be extrapolated simply from estimates made for developed countries. The impacts of extreme temperatures are also uncertain. Current damage estimates are generally calibrated for warming of less than 3 °C (ref. 6). Yet without mitigation, the IPCC projects that we could see warming in excess of 4 °C by the end of the century. Such conditions would

be beyond human experience. If warming continues unchecked into the twenty-second century, it could render parts of the planet effectively uninhabitable during the hottest days of the summer, with consequences that would be challenging to monetize[10].

The models should be revised more frequently to accommodate scientific developments. Researchers commonly test model sensitivity to new parameters. But the structure and in some cases the calibration of the damage models is stuck in the 1990s, when the original versions were created, owing to a lack of funding.

IPCC reports help to set the research agenda on climate. The release of the Fifth Assessment Report reminds us of the progress so far. It is important to ensure that the sixth assessment takes a substantive step forward. By facilitating efforts to refine estimates of the social cost of carbon, the IPCC will be performing its most important function: informing the global political conversation about how best to address the looming threat of climate change. ∎

Richard L. Revesz and **Peter H. Howard** *are at the New York University School of Law, New York, USA.* **Kenneth Arrow** *and* **Lawrence H. Goulder** *are in the Department of Economics, Stanford University, Stanford, California, USA.* **Robert E. Kopp** *is in the Department of Earth and Planetary Sciences and Rutgers Energy Institute, Rutgers University, New Brunswick, New Jersey, USA.* **Michael A. Livermore** *is at the University of Virginia School of Law, Charlottesville, Virginia, USA.* **Michael Oppenheimer** *is at the Woodrow Wilson School of Public and International Affairs and in the Department of Geosciences, Princeton University, Princeton, New Jersey, USA.* **Thomas Sterner** *is in the Department of Economics, University of Gothenburg, Gothenburg, Sweden.* e-mail: richard.revesz@nyu.edu

1. Interagency Working Group on Social Cost of Carbon. *Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis* (US Government, 2013); available at http://go.nature.com/vzpkkb.
2. Interagency Working Group on Social Cost of Carbon. *Social Cost of Carbon for Regulatory Impact Analysis* (US Government, 2010); available at http://go.nature.com/uqhrgh.
3. Kopp, R. E. & Mignone, B. K. *Economics* **6,** 2012–2015 (2012).
4. Schlenker, W. & Roberts, M. J. *Proc. Natl Acad. Sci. USA* **106,** 15594–15598 (2009).
5. Fankhauser, S. & Tol, R. S. J. *Resour. Energy Econ.* **27,** 1–17 (2005).
6. Tol, R. S. J. *Annu. Rev. Resour. Econ.* **3,** 419–443 (2011).
7. Hsiang, S. M., Burke, M. & Miguel, E. *Science* **341,** 1235367 (2013).
8. Sterner, T. & Persson, U. M. *Rev. Environ. Econ. Pol.* **2,** 61–76 (2008).
9. Weitzman, M. L. *Environ. Econ. Mgmt* **36,** 201–208 (1998).
10. Sherwood, S. C. & Huber, M. *Proc. Natl Acad. Sci. USA* **107,** 9552–9555 (2010).

SPRING BOOKS

BIOTECHNOLOGY

Recombinant gold

Nathaniel Comfort applauds a nuanced history of genetic engineering's early years.

In 1969, the molecular biologist Gunther Stent published one of the most spectacularly inaccurate predictions in the history of modern science. In *The Coming of the Golden Age: A View of the End of Progress* (Natural History Press), he stated his belief that molecular genetics — which had only really been a science for 15 years — had peaked. The "golden age," he wrote, would be one of modest discovery and waning public interest in science. That year, Jonathan Beckwith isolated the first gene. In 1970, Hamilton Smith found the first

site-specific restriction enzyme, which his colleague Daniel Nathans developed into a tool for cutting and pasting DNA. Then, in 1972, Paul Berg spliced a bacterial gene into a virus. With the ability to engineer genes, molecular genetics began in earnest. Never mind the Age of Aquarius; this was the age of recombinant DNA.

In *Gene Jockeys*, the biologist and science historian Nicolas Rasmussen delicately unravels the tangled fibres of discovery, entrepreneurship and lab life in the first decades of genetic engineering. Moving drug

by drug through a series of case studies, he depicts a world of scientists in transition from curiosity-driven research to commercial biotechnology. He then shades in the outlines with the internal contradictions, the rationalizations, and the mixed feelings with which biotech pioneers moved towards a profit model of science. Rasmussen's research is dogged and creative, his analysis perceptive and nuanced. Although his inner gene jockey shows through in a wealth of technical detail, any scientifically literate reader will find an engaging, ultimately

the recombinant-DNA technique. Cohen's institution, Stanford University, quickly filed for a patent on it. The following year, anxious about having regulations imposed from without, Berg and other recombinistas self-imposed a moratorium on genetic-engineering experiments. Boyer, however, neatly sidestepped it, performing experiments including the insertion of a frog gene into a bacterium. At Asilomar, California, in 1975, molecular biologists agreed to lift the moratorium. The following year, Boyer's fledgling company, Genentech, drafted its first business plan and quickly settled on cloning human insulin as its first commercial project. Thus, Rasmussen argues, biologists set the intellectual values, technical proficiency, ethical standards and regulatory policy during the early development of biotech. An overheated investment climate and frenzied marketing took it from there.

Curiosity-driven science continued to shape both the marketplace and the legal environment in which biotech could thrive into the 1980s. This led to some strange dance partners. Tracing the development of the cancer drug α-interferon, Rasmussen examines the "pas de trois" of the commercially sponsored clinical trial: government regulators requiring private companies to pay university researchers to test their own drugs for efficacy. Erythropoeitin (Epo) — a hormone that stimulates red blood cell production — was the first true blockbuster drug to come out of the new genetics. But Epo was a mixed blessing, Rasmussen writes. By moulding the biotech reward system to favour expensive, high-tech drugs, it primarily benefited wealthy patients, further stratifying US medical care. Rasmussen acknowledges that recombinant interferon and Epo were significant medical contributions, but insists that they were also grossly oversold.

With such triumphs, the scales tipped: commercial interests began to drive the science. As Rasmussen shows, recombinant tissue-type plasminogen activator, or tPA, a clot-dissolver that could be used to mitigate a heart attack in progress, was hailed as the "billion-dollar protein". The stakes were now too high to sustain the lofty, lefty ideals of 'pure' communal science. Researchers became entrepreneurs. Discoveries became intellectual property. Patenting preceded publishing. The science never disappeared, but business values increasingly trumped it. Rasmussen is wistful for the early days of

biotech's golden age, when science drove business rather than the other way around.

The parallels between biotech and the contemporaneous (and neighbouring) rise of computing are relevant, given the degree to which genomics and information science have merged. Rasmussen does not provide this context, which would have situated the history better and helped to connect it to the present day. The contrasts between computing and biotech are as interesting as the continuities. Unlike the nerds of Silicon Valley, who started up their companies in garages, the bio-geeks started theirs in well-stocked, mainly government-funded university labs. But, as in IT, the dream became having your little company bought by one of the big corporations. By the 1990s, both industries had become less freewheeling and curiosity-driven, and more privatized and gold-directed.

This trajectory, of course, tracks society's. The Human Genome Project began around the time that the Berlin Wall came down and the Soviet Union collapsed. Soon, a collective mythology emerged — particularly in the United States — that capitalist individualism had 'won'. The resulting shift towards privatization is hitting every sector, but biomedicine is a sharp lens through which to view it. Molecular biology still has its idealists, but their ponytails are greying. Today's graduate students want training in economics, marketing and management, even MBAs. You can now go into biology for the money.

Perhaps Stent wasn't as far off as we thought. Certainly, he missed big by predicting the end of progress. But he presciently foreshadowed, if not the coming of the golden age, the coming of the age of gold. ■

Gene Jockeys: Life Science and the Rise of Biotech Enterprise
NICOLAS RASMUSSEN
Johns Hopkins University Press: 2014.

> GRADUATE STUDENTS WANT TRAINING IN ECONOMICS, MARKETING AND **MANAGEMENT.** YOU CAN NOW GO INTO BIOLOGY FOR **THE MONEY.**

elegiac tale of lost innocence, as researchers struggle with the angel of the search for truth on one shoulder, and the devil of wealth and fame on the other.

Imagine a gold rush sparked not by the discovery of gold, but by brainy miners lured by the problems of extraction, speculating that their methods could be applied to mining precious metals; initially, the gold itself is a proof of concept. In Rasmussen's view, the first recombinant drugs were no mother lode. They met no demand in the medical marketplace. Rather, development was driven mainly by scientific curiosity. But the culture of academic science created an environment favourable to commercial development of genetically engineered drugs.

For example, in 1973, Stanley Cohen and Herbert Boyer developed an elegant variation on

Nathaniel Comfort *is professor of the history of medicine at Johns Hopkins University in Baltimore, Maryland. His most recent book is* The Science of Human Perfection: How Genes Became the Heart of American Medicine.
e-mail: comfort@jhmi.edu

ILLUSTRATION BY MARTIN O'NEILL; ANT PHOTO: ANTHONY BANNISTER/NHPA/PHOTOSHOT

ECOLOGY

Wilson in Africa

Stuart Pimm enjoys a fellow naturalist's first visit to sub-Saharan Africa, and the global lessons drawn from it.

Speeding northwards, I luxuriated in the smooth South African highway: soon we would cross into Mozambique, where I could expect a rutted dirt road. But there wasn't even that. Beyond the border-post's single hut, a braid of narrow, indistinct tracks headed in all directions. We picked the travelled paths; there might be land mines along unbeaten ones. Mozambique had suffered a brutal war of independence and subsequent upheavals between 1964 and 1992. An estimated one million Mozambicans died, wildlife was slaughtered and forests burned. So why were we there?

For the same reason as Edward O. Wilson — the region's exceptional biodiversity. As he explains in *A Window on Eternity*, "Anywhere I am in the world I love it when the air is warm and moist and heat bounces off the sunlit earth, and insects swarm in the air and alight on flowers." In the book, Wilson explores and revels in Mozambique's Gorongosa National Park, where the wet heat makes for a riot of what the eminent biologist calls "the little things that run the world" — insects and other invertebrates. At the heart of the park is Mount Gorongosa, which, at

1,863 metres high, catches enough moisture from Indian Ocean winds to support a dripping rainforest. In East Africa, you can move from dry grasslands to savannah woodlands to montane rainforest in the space of a few kilometres, with each habitat harbouring unique species.

Unexpectedly, Wilson is new to sub-Saharan Africa: the book is a chronicle of his first visit there in 2011. His responses recall Darwin's enthusiasm on first encountering the Brazilian tropics, marvelling breathlessly at one fascinating species after another.

Inevitably, ants — his speciality — take centre stage. Their social systems led to *Sociobiology: The New Synthesis*, his landmark 1975 study on animal societies. Gorongosa's ants are new to Wilson, however. He picks up a Matabele ant (*Pachycondyla analis*), which "gnashed its mandibles impressively, then … thrust a long sting into the flesh of my index finger". (He rates the pain as slightly below that of a hornet's sting.) These are the ant equivalent of "sappers and light cavalry" — heavily armoured and specialized for raiding termite communities that build hardened mounds of mud that can reach the size of a bus. The termites have their own soldiers, but the ants overwhelm them by sheer force of numbers and ferocity.

Wilson also delights in driver ants, whose colonies can number 20 million workers. They are blind, but sensitive to smell and movement. Their leaderless swarms engulf the ground and low vegetation, seizing almost any live animal in their path. Useful things, driver ants: nothing quite like a visit from them to clean out vermin-infested tropical homes.

There is a world of sound, too, although it is mostly hidden. Entomologist and photographer Piotr Naskrecki, who accompanied Wilson, gathers the audio. He brings special equipment to record
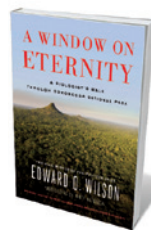
A Window on Eternity: A Biologist's Walk Through Gorongosa National Park

EDWARD O. WILSON
Simon & Schuster: 2014.
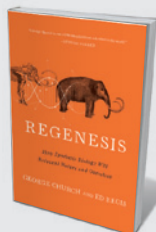
Regenesis: How Synthetic Biology Will Reinvent Nature and Ourselves
George Church and Ed Regis (Basic Books, 2014)
In this visionary account of humankind's soaring capabilities in bioengineering, geneticist George Church and science writer Ed Regis redraw the frontier of synthetic biology — and, potentially, some ethical boundaries. Starting with 'simple' bioplastic cups made entirely from plants, they go on to discuss genomic alterations that could spawn virus-resistant humans and resurrect Neanderthals. This is the first book to be translated into a DNA sequence (see go.nature.com/gcgfqa).

sounds of up to 250 kilohertz, well above our human limit of 20 kHz. "The unaided ears of a human walking through the forest at night are assaulted by a riot of unheard katydid cross-talk," Wilson recounts. We humans — and, presumably, potential predators — hear warning sounds, but not other, still vital communications.

A Window on Eternity revels in biodiversity and nature's inventiveness. Wilson damns "the corporate priesthood" that views "restructuring … Earth to accommodate vast numbers of people" as progress. There may be those to whom species do not matter, to whom extinction is an abstraction. To Wilson, species are our "phylogenetic kind" and individual species matter to him. He indicts those for whom 'Anthropocene' is a term that carries the political baggage of acquiescence to human domination of landscapes. The world cannot dwindle into a vast garden, he urges. To him, wildlands are "our birthplace"; a further "slide into extinction will turn the Anthropocene into the Eremocene, the Age of Loneliness".

His choosing Gorongosa is surely no accident. In common with much of Mozambique, it lost almost all of its large animals during its wars: by 2001, buffalo had dropped from 13,000 to 15; wildebeest from 6,400 to 1; and hyenas and rhinos had become locally extinct. Entrepreneur and philanthropist Greg Carr drove across the area in 2004, going days without seeing large mammals. He initiated the Gorongosa Restoration Project to plant trees, reintroduce large mammals, and create a tourist centre to make the park self-sustaining. Wilson plants his defiant flag defending biodiversity in a place once so brutally despoiled that its recovery is truly momentous. ∎

Stuart Pimm *is professor of conservation at the Nicholas School of the Environment, Duke University, Durham, North Carolina, USA, and author of* The World According to Pimm: A Scientist Audits the Earth. *e-mail: stuartpimm@me.com*

A strained relationship

Scott Barrett examines a study probing the nexus between climate change and energy.

Michael Grubb's provocative *Planetary Economics* claims to be about the "grand challenges of energy and environment", but is really about the relationship between energy and climate change. Grubb briefly notes the scale of the problem: for atmospheric levels of greenhouse gases to be stabilized, net emissions must fall to zero. His focus, however, is on reducing energy consumption and carbon emissions, irrespective of the need to meet any particular target for greenhouse-gas concentrations. It is unclear how much emissions would fall if the book's ideas were actually implemented.

Written with input from fellow climate-policy researchers Jean-Charles Hourcade and Karsten Neuhoff, this is a long and at times repetitive book; but there is something interesting on every page. It reflects a wealth of accumulated wisdom: Grubb has engaged with these issues for more than two decades.

He is critical of dominant theories, such as the assumption that economic agents are rational and optimize every decision. He says that this approach fails to capture the complexities, overestimating the costs of reducing emissions and underestimating the benefits. He rejects the cost–benefit framing and its estimates of the "social cost of carbon" — a concept that puts a monetary value on the damage associated with a 1-tonne increase in carbon dioxide emissions. Instead, Grubb seems to endorse the political target of keeping global temperature rise below 2 °C, as agreed at the 2009 meeting of the United Nations Framework Convention on Climate Change (UNFCCC). Yet the economist Nicholas Stern has supported the same target, in part with reference to the social cost of carbon. It is true that greenhouse-gas concentrations have continued to creep up, threatening any possibility of achieving the target. However, there is no evidence that economic theories are to blame for the failure
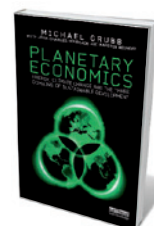
Planetary Economics: Energy, Climate Change and the Three Domains of Sustainable Development
MICHAEL GRUBB WITH JEAN-CHARLES HOURCADE AND KARSTEN NEUHOFF
Routledge: 2014.

to cut emissions.

What is the solution? Grubb's thesis is that reducing energy consumption and its associated emissions requires policy changes in three domains: satisficing, optimizing and transforming. In satisficing, people and firms overlook cost-effective ways to save energy, such as investing in insulation that promises a reduction in energy bills. Policy can help by nudging an economy closer to the energy-use efficiency frontier, for example by introducing standards for appliances, production processes, buildings and cars. In the second domain, optimizing, people and firms respond to price signals; this ushers in an argument in favour of carbon taxes and cap-and-trade schemes, which allow participants to trade emissions allowances under an overall cap. Transforming involves revolutionizing technology through innovation and investment in infrastructure such as improved electricity transmission.

Grubb argues that advancement demands progress in all of these mutually reinforcing domains. This argument is compelling, but the real questions are how far policy should go in each domain, and precisely how such policy should be devised. How should decisions be made about setting standards, designing cap-and-trade schemes and choosing strategic investments, if not through a cost–benefit rule? How should a carbon tax be chosen if not with reference to the social cost of carbon? Grubb challenges the idea that the price of carbon should be the same everywhere, ▶
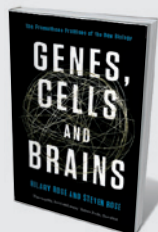
Genes, Cells and Brains: The Promethean Promises of the New Biology
Hilary Rose and Steven Rose (Verso Books, 2014)
In this exposé of clashes between society and science, sociologist Hilary Rose and neurobiologist Steven Rose lambast multibillion-dollar biotech research, showing how the Human Genome Project, for instance, has not found disease-triggering genes.

Forecast
Mark Buchanan (Bloomsbury, 2014)
Disassembling the "marvellous machine" of the free market, physicist Mark Buchanan analyses the tempestuous global economy. Principles such as positive feedback loops and fluid dynamics explain the market's natural instability and inform ways to weather future fiscal storms.

▶ supporting instead a "base level" price for developing countries, with others setting higher prices. He backs this in part by appealing to basic welfare economics. But that, as he notes, assumes that financial transfers from rich to poor countries are infeasible, which sits oddly with the fact that rich countries have pledged billions of dollars to the UNFCCC's Green Climate Fund to help poorer nations to mitigate emissions and adapt to climate impacts.

The book's greatest weakness is its lack of an overarching framework. The introduction acknowledges the importance of global collective action, but states that the problem is beyond the book's scope. Grubb concludes that the "next phase of the global effort … is a question of investment and returns". But which nations are to make the investments, which to earn the returns? Countries care about the answers. That is why so much effort has gone into climate negotiations.

Such answers will not be central to how countries address most of the issues in Grubb's first domain. Measures to increase energy efficiency must satisfy mainly domestic criteria, although international trade links will be important for setting technology standards. However, global questions and answers are key to the other domains. They will matter when a country sets a carbon price. They will matter in relation to the investment that countries are willing to give to decarbonization, not least because such investments will pay off only if the carbon price is high.

Despite the criticisms, the book's thesis is relevant to current climate negotiations, which seem to be focusing on what countries are willing to do individually: a kind of enhanced 'business as usual' approach. But for the immense scale of action needed to stabilize concentrations of greenhouse gases — even at a level allowing global temperatures to rise by more than 2 °C — collective action is essential. ∎

Scott Barrett *is Lenfest-Earth Institute Professor of Natural Resource Economics at Columbia University in New York, and co-editor of the forthcoming* Environment and Development Economics. *e-mail: sb3116@columbia.edu*

MEDICINE

Outside the fold

Giovanna Mallucci assesses the autobiography of Stanley Prusiner, the discoverer of prions.

In his autobiographical book, *Madness and Memory*, Stanley Prusiner charts his journey to winning the 1997 Nobel Prize for Physiology or Medicine, 30 years after his first laboratory posting as a fourth-year medical student — in Sweden, coincidentally. As the subtitle states, the story hinges on Prusiner's discovery of prions, "a new biological principle of disease". It is a remarkable tale: for the scientific discovery at its centre; for its recording of the extraordinary resistance the idea engendered; and for the sheer unfettered animosity, both personal and professional, directed at him throughout from scientists and the media.

Prusiner introduces us early on to how, in 1972, his imagination was caught by the then-unknown infectious agent causing scrapie, a degenerative brain disease of sheep. Then, in the first year of his neurology residency, he was caring for a patient with Creutzfeldt–Jakob disease (CJD), a scrapie-like neurodegenerative disorder we now know to be a human prion disease. What follows is a chronicle of his voyage to isolate and characterize the scrapie agent during the 1970s, up
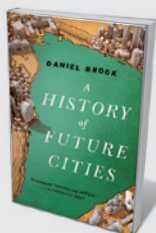
A History of Future Cities
Daniel Brook (W. W. Norton, 2014)
As urban innovation soars in skyscraper-studded Dubai, Daniel Brook looks to the original instant cities, where Western architecture invaded Eastern streets. St Petersburg, Shanghai and Mumbai, he argues, traded culture for economic power. (See Mike Davis's review: *Nature* **494,** 427–428; 2013.)
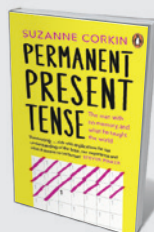
Permanent Present Tense
Suzanne Corkin (Penguin, 2014)
Neuroscientist Suzanne Corkin worked with Henry Molaison, who had irreparable memory loss, for decades. Here she documents discoveries such as the hippocampus's role in classifying memories. (See Douwe Draaisma's review: *Nature* **497,** 313–314; 2013.)
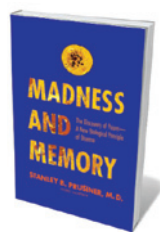
Madness and Memory: The Discovery of Prions — A New Biological Principle of Disease
STANLEY B. PRUSINER
Yale University Press: 2014.

to his demonstration in 1981 that the agent is a protein, devoid of nucleic acid and capable of replicating itself and transmitting disease.

Details of scientific milestones are interwoven with Prusiner's accounts of personal interactions with colleagues at each stage; descriptions of dramatic epidemics of prion diseases across the globe punctuate the story. These include the phenomenon of kuru, a scrapie-like disease first described by Carleton Gajdusek in the 1950s, found among the Fore people of Papua New Guinea who practised ritualistic cannibalism; and the late 1980s disaster of bovine spongiform encephalopathy (BSE) in UK cattle, which spread to humans as variant CJD. Altogether, the story highlights the extraordinary scientific, medical and political climate of this period. There are struggles for funding, Prusiner's battle for tenure at the University of California, San Francisco, and the fight for scientific recognition at every stage of his career.

The idea of an infectious protein was first mooted in London in the 1960s by the mathematician John Stanley Griffith, and separately by the radiation biologist Tikvah Alper. Both published papers in *Nature* in 1967: 'Self-replication and scrapie' (J. S. Griffith *Nature* **215**, 1043–1044; 1967) and 'Does the agent of scrapie replicate without nucleic acid?' (T. Alper *et al. Nature* **214**, 764–766; 1967). But it was Prusiner who pursued the idea and Prusiner who proved it, with a determination that earned him both admiration (including the Lasker and Nobel prizes) and blind hostility.

Prusiner's proof of the revolutionary concept that infectivity in these disorders proceeds without nucleic acids, and his coinage of the term prion — for proteinaceous infectious particle (S. B. Prusiner *Science* **216**, 136–144; 1982) — engendered disbelief, anger, denial and widespread refusal from a hostile scientific community, including many virologists. This antipathy persisted right up to the awarding of his Nobel prize. A 1997 article in *The New Yorker* asked, "But what if he's wrong?"; one in *Slate* was entitled 'Nobel Gas: Sure, Stanley Prusiner deserves a prize — for his persistence, not for his prions.' That Prusiner withstood these public attacks in his moments of greatest recognition is laudable, and gives some insight into his tenacity.

Scientifically, the story progresses from descriptions of labs full of antediluvian centrifuges with custom-built safety cabins around them (for isolation of the infectious agent from infected hamster brains) to the elegant transgenic mouse experiments of the prion era and the concept of misfolded endogenous proteins associated with disease. The story is testament to the staggering intellect and courage involved in one of the most exciting discoveries since the DNA double helix.

But for all its drama, this extraordinary scientific story is not always an easy read. Prusiner's description of scientific rivalries inevitably reflects his version of events. His relationship with Gajdusek, whom he had visited in Papua New Guinea, ends bitterly, with arguments over Gajdusek's claims to the prion concept. Prusiner's collaboration with Swiss molecular biologist Charles Weissmann began brilliantly but became strained. Weissmann discovered that prion protein is encoded by a host gene — a finding as important to the prion story as the infectious protein itself. Prusiner's account of that collaboration's end, complete with transcripts of letters he sent to Weissmann, is an uncomfortable read. 'Scientific Interludes' explaining concepts and technicalities in some detail are interspersed through the text. There are other inserts: technical, graphical and bibliographical. Many are fascinating, but they tend to interrupt the flow of the compelling narrative.

The concluding chapter's meditation on the prion concept in widespread neurodegenerative disorders such as Alzheimer's is highly topical, albeit inordinately specialized in its detail. The book ends with an exhortation to the US government to fund research into cures for neurodegenerative disorders, diseases that Prusiner claims have "been ignored". He will be glad that last December's G8 Dementia Summit means that the world is heeding his call and taking this issue seriously at last. ∎

> **PRUSINER'S PROOF ENGENDERED DISBELIEF, ANGER, DENIAL AND WIDESPREAD REFUSAL.**

Giovanna Mallucci *is professor of neuroscience at the MRC Toxicology Unit, Leicester, UK, and honorary consultant neurologist at Addenbrooke's Hospital, Cambridge, UK. She works on neurodegeneration, focusing on prion diseases.*
e-mail: grm7@le.ac.uk

Time Reborn: From the Crisis in Physics to the Future of the Universe
Lee Smolin (Mariner, 2014)
Scientists unscrambling the fundamentals of the Universe dub time an illusion. Theoretical physicist Lee Smolin resurrects the concept as a constant around which other universal laws evolve. (See Pedro Ferreira's review: *Nature* **496,** 430–431; 2013.)

Jane Austen, Game Theorist
Michael Chwe (Princeton Univ. Press, 2014)
Using Jane Austen's novels and craftiest characters, such as George Wickham in *Pride and Prejudice*, Michael Chwe proves that game theory — mathematics-based strategizing — has been harnessed for social and emotional advancement as well as for military victory.

DRUGS

Gut response

Maryn McKenna finds much to digest in a warning about the demise of our bodily bacteria.

Last year, two public-health-agency chiefs chose dramatic language to alert their nations to a menacing health problem — a rise in the spread and severity of bacterial resistance to antibiotics. UK chief medical officer Sally Davies called it a "catastrophic threat"; Thomas Frieden, director of the US Centers for Disease Control and Prevention, spoke of a "nightmare". They were flagging the emergence of an almost pan-resistant bacterium, carbapenem-resistant Enterobacteriaceae. This is the latest in a series of tough-to-treat organisms — the result of overuse of antibiotics since the 1940s.

In *Missing Microbes*, Martin Blaser sounds a related alarm. He patiently and thoroughly builds a compelling case that the threat of antibiotic overuse goes far beyond resistant infections. Antibiotics, he warns, are destroying the benign bacteria that are crucial to the functioning of human bodies, and this trend is contributing to health problems from obesity to diabetes and bowel disease.

Antibiotic resistance can be devastating for patients, but it has until recent years received scant attention from policy-makers, as Blaser knows. A physician and microbiologist, he has treated many people with resistant infections. And as past president of the Infectious Diseases Society of America, he pressed the US government to curb antibiotic overuse and encourage drug development.

Blaser, who is also director of New York University's Human Microbiome Project, investigates the personal bacterial ecosystems that allow us to absorb nutrients and develop immunity. He has long studied *Helicobacter pylori*, a bacterium that thrives in the stomach and can cause ulcers and stomach cancers. Using mouse experiments and epidemiological data, he has shown that *H. pylori* may also be associated with a reduced incidence of asthma, allergies and severe reflux disease. That should concern us, he writes, because worldwide, the rate of *H. pylori* infection is going down.

The make-up of human microbiomes is shifting, with diversity declining and keystone species disappearing. Blaser blames these changes on innovations that impede bacterial attempts to set up shop inside us. Examples include germ-killing hand sanitizers, and Caesarean sections that rob newborns of the bacterial kick-start that they usually get by passing through the birth canal. In the United States alone, one-third of all births are now Caesareans.

Blaser is strongest, and most provocative, when he questions a practice that has become routine in much of the industrialized world: feeding small doses of antibiotics to meat animals as growth promoters. The early discovery that antibiotics work as fattening agents gave birth to the entire structure of modern concentrated meat-raising. By saturating the environment with antibiotic residues, Blaser argues, we have effectively recreated that weight-gain programme in humans — and the result has been the seemingly unstoppable increase in obesity, especially in children.

Can the trend of inadvertently destroying our microbiomes be reversed? Blaser is sceptical. The public has been indifferent to warnings about resistance since antibiotic use began: penicillin discoverer Alexander Fleming cautioned in his 1945 Nobel prize acceptance speech that using the drugs carelessly would undermine their power to treat

---
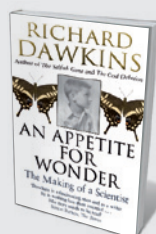
An Appetite for Wonder: The Making of a Scientist
Richard Dawkins (Black Swan, 2014)
From childhood in Africa to enlightenment in Oxford, Richard Dawkins chronicles his life before *The Selfish Gene* (1976) and the nature and nurture of his science obsession. (See Eugenie Scott's review: *Nature* **501,** 163; 2013.)

Gulp: Adventures on the Alimentary Canal
Mary Roach (W. W. Norton, 2014)
Journalist Mary Roach offers a consummate tour of the digestive system. Investigating dog-food-testing facilities, brain-munching Inuits and optimum mastication techniques, Roach maintains a balanced diet of wit and reliable gastro-wisdom. (See David Katz's review: *Nature* **495,** 446; 2013.)
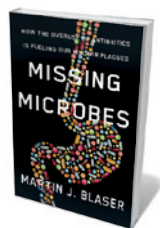
Missing Microbes: How the Overuse of Antibiotics is Fueling Our Modern Plagues
MARTIN J. BLASER
Henry Holt: 2014.

lethal diseases. Antibiotics — like Caesarean sections and hand sanitizers — make modern life easier. Blaser would say that we have ignored Fleming's warning: we have consistently chosen convenient over smart and safe.

The first step, Blaser says, is to roll back antibiotic overuse in agriculture and medicine. After that, he suggests, we should bend our appetite for innovation towards finding ways to repair the microbiome damage that the wonder drugs have done.

Blaser foresees the development of microbial supplements — a more sophisticated version of the faecal transplants already being used in some quarters to combat *Clostridium difficile* infections — that could restore microbial communities devastated by antibiotics. The regulation of faecal transplants has confounded the US Food and Drug Administration (M. B. Smith *et al. Nature* **506,** 290–291; 2014), but many patients — along with academic researchers in Europe and Australia — have taken to them intuitively and enthusiastically. It seems likely that Blaser's concept of personal, protective microbial cocktails would also find support.

It is urgent that we take these steps soon. *Missing Microbes* explains that our ancient microbiome is akin to an essential organ; we unthinkingly excised it, and only now are waking up to the implications. Changes to it come with costs, Blaser warns, "but we are only just beginning to recognize them. They will escalate." ∎

Maryn McKenna *is the author of* Superbug *and* Beating Back the Devil, *and a senior fellow at the Schuster Institute for Investigative Journalism at Brandeis University. She lives in Atlanta, Georgia.*
e-mail: mmckenna@mindspring.com

EDUCATION

Digital lessons learned

Robert Lue enjoys a deft study of online pedagogy.

Massive open online courses, or MOOCs, have generated unprecedented debate over their educational value and impact on the university ecosystem in the past three years. Deploying videos, online assessments and discussion forums to teach potentially 100,000 students at a time, MOOCs have taken off with the advent of for-profit companies such as Coursera in Mountain View, California, and non-profit efforts including edX, founded by Harvard University and the Massachusetts Institute of Technology (MIT) in Cambridge. As the faculty director of HarvardX, the Harvard initiative involved with edX, I have observed the pendulum of media opinion swing between extremes.

Initially MOOCs were lauded as the ultimate way to deliver educational content: they could reach the farthest corners of the globe and cut costs to promote sustainability. Now, views have reversed: the schemes seem to be reaching mainly those who already have a degree, and concerns are emerging that they will threaten faculty jobs in some institutions. The reality will almost certainly lie between.

In her provocative book *The War on Learning*, Elizabeth Losh fires several warning shots across the bows of online education. Along with MOOCs, she takes on active learning methods such as real-time online polling. She provides welcome context on both the recent history of distance education, covering efforts such as the OpenCourseWare movement launched by MIT in 2002, and the surprising contribution of applications such as iTunes. Her book is a timely exploration of the sometimes daunting but often rewarding faculty

and institutional experience of teaching with technology; it touches on students' experience more lightly.

Losh is critical of both irrational exuberance over the reach of technology, and panic that it has destroyed the classroom. She presents anecdotes documenting instructors' successes and failures, including her own misadventures using Twitter to engage a large lecture class in the Culture, Art and Technology programme at the University of California, San Diego. It became a platform for humorous expressions of student inattention and class-wide pranks. Indeed, not all modes of digital engagement are suited to education: anonymous mass action can result solely in mischief. Losh seeks a pedagogical silver lining by connecting classroom Twitter with new ways for students to engage with both content and the learning experience.

> ## NOT ALL MODES OF DIGITAL ENGAGEMENT ARE SUITED TO EDUCATION.

Losh's rich selection of anecdotes swerves between triumphs and tragedies — so much so that real success becomes difficult to recognize. She seems to revel in this ambiguity, perhaps seeking to underscore just how much remains unproven and unsettled in a field still finding its way. Is the compelling but carefully scripted drama of a TED Talk a model for engaging students online — or an opportunity for mutual narcissism, whereby the speaker tells the audience exactly what they want to hear? Does the professor lauded for his or her classroom oration promote self-interest above learning when their online reach goes global?

Healthy scepticism is Losh's dominant tone, especially in her discussion of 'gamification' — the use of video-game ▶
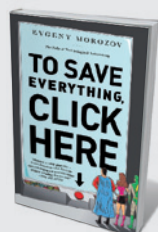
To Save Everything, Click Here: The Folly of Technological Solutionism
Evgeny Morozov (PublicAffairs, 2014)
Can a toolbox of techno-fixes really solve climate change, disease and crime? Social theorist Evgeny Morozov rebels against a technocratic Utopia and critiques the ideology of computerized cure-alls. (See Nicholas Carr's review: *Nature* **495,** 45; 2013.)

Brilliant Blunders: From Darwin to Einstein
Mario Livio (Simon & Schuster, 2014)
Astrophysicist Mario Livio reveals the epic errors of famous scientists that yielded glorious breakthroughs. When Einstein misconstrued universal equilibrium, his equations led to the discovery of the expanding cosmos. (See Mario Livio's Comment: *Nature* **497,** 309–310; 2013.)

ILLUSTRATION BY MARTIN O'NEILL

▶ approaches, such as the modelling in the virtual world *Second Life*. In practice, as she shows, developing the expertise to reach the next level of a games environment is not necessarily the same as learning: obeying a set of rules does not automatically provide insight into underlying concepts or their creative application. Ironically, some approaches lose the very fun that motivates: the Shakespeare game *Arden*, for example, fails to capture the emotional element of the plays. In story after story of instructional technology failing to reach its transformative potential, Losh reveals that by focusing on the broadcasting of information, usually through ever more nifty gadgets, instructors and institutions risk losing sight of education as an iterative process that is based on dialogue and group reflection.

Losh concludes with suggestions for how universities might better harness technology to serve teaching. She argues that the physical classroom should be revived, because "learning benefits from embodied performances of knowledge", and using these spaces for "joyful" educational activities can offer profoundly engaging experiences.

Hackathons, for instance, bring students together for intense periods of participatory learning around common challenges, such as writing code. Likewise, Losh warns against blindly valuing innovations above clear pedagogical purpose, or eschewing tried-and-tested tools such as e-mail.

Losh falls victim to some romanticism in her recommendation that faculty members share their own research tools (say, for mapping or imaging) as part of their teaching, rather than simply give students gadgets such as clickers. This is a laudable suggestion that builds bridges between scholarship and teaching while exposing students to digital tools that have real currency in the world of research. However, such an approach could leave behind students who are not in the upper echelons of the class.
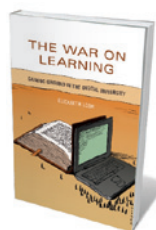
The War on Learning: Gaining Ground in the Digital University
ELIZABETH LOSH
MIT Press: 2014

Simpler, well-executed pedagogical tools might better serve a wider audience. Merely touching genuine research tools does not ensure a good learning experience, especially if there is nothing to prepare you for it.

The War on Learning is a welcome addition to the debate and is well informed by the literature in this area, as well as cognitive psychology. That said, by largely restricting her focus to instructor and institution, Losh misses an opportunity to address the increasingly complex ecosystem of higher education, including the instructional designers, education researchers and technologists who so often work side by side with faculty members on the crafting of courses on campus and online. As this web of relationships evolves, it will redefine how instructors and institutions sustain innovative teaching through ever deeper engagement with technology. ∎

Robert A. Lue *is faculty director of HarvardX and the Derek Bok Center for Teaching and Learning at Harvard University in Cambridge, Massachusetts.*
e-mail: lue@harvard.edu
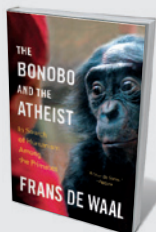
The Bonobo and the Atheist: In Search of Humanism Among the Primates
Frans de Waal (W. W. Norton, 2014)
Primatologist Frans de Waal probes the roots of "human" emotions and morality in the bonobo, whose behaviours include empathy, altruism and even remorse. (See Christopher Boehm's review: *Nature* **495,** 312; 2013.)

The Recursive Mind: The Origins of Human Language, Thought, and Civilization
Michael C. Corballis (Princeton Univ. Press, 2014)
Does language alone make us human? Psychologist Michael Corballis takes a fresh look at the origins of speech, and suggests that our ability to embed thoughts within other thoughts — recursion — gives our species the edge.

ENERGY

The new oil era

Chris Nelder relishes a lively history of fracking that delves into the complexities.

Despite its title, *The Boom* is no sales pitch for fracking. *Wall Street Journal* energy reporter Russell Gold has produced a thoughtful piece of journalism, exploring the complex landscape of drilling, finance and politics that brought a gusher of oil and gas to a country convinced that its hydrocarbon heyday was over. Gold offers no pat answers to the challenges that this new abundance poses, but reminds us starkly of "unforeseen costs and necessary evils".

Skilfully interweaving hard data about US energy with an engaging narrative, Gold covers previous oil booms in Texas, Oklahoma, the Niger Delta, the Bakken shale in North Dakota (which saw its first oil well drilled in 1953), and Pennsylvania. He also relates the biography of businessman George Mitchell, 'the father of fracking'; the fundamentals of petroleum geology; and the long evolution of technology for oil and gas production. Much of *The Boom* focuses on Aubrey McClendon, founder of the Chesapeake Energy company. Describing him as "part pied piper, part early adopter, and part rapacious capitalist", Gold duly credits McClendon for seeing the potential in shale gas and driving its production, but gives equal weight to his rise and downfall.

As a lens on the conflicts between ethics and raw need that characterize today's energy industry, Gold uses a personal experience. His parents had co-owned land in Pennsylvania since the 1970s. In 2009, Chesapeake Energy was snapping up leases there to drill for natural gas in the Marcellus shale formation underneath. As he advised his parents on Chesapeake's US$400,000 offer, Gold was torn between their desire to prevent groundwater contamination, and his certainty that fracking would proceed around them, driven by the need for domestic gas, income and jobs. As one farmer told him, local exploitation of forests and coal had been unsustainable:

"We can't do that three times in a row ... If we don't do this right, what the hell have we done?" Despite the doubts, the lease was signed.

Along with his hard-hitting analysis of the disruptive nature of fracking, Gold offers deep reportage on two previously untold stories. The first concerns the close relationship between McClendon and his best friend since university, Ralph Eads. Starting in 2004, Gold says, McClendon talked to Eads regularly about how to round up more capital. Eads "set out to create a new financial ecosystem to find money to drill shale", pitching Chesapeake's opportunities to institutional investors globally. The company raised $33.7 billion, with Eads serving as financial adviser for nearly all of the deals. Chesapeake also bet heavily on the future price of gas.

By 2008, Gold writes, "McClendon and Chesapeake were whales in the futures market". But as competitors arrived, a gas glut developed and prices plummeted. "With its heavy debt and drilling commitments, this price decline was a recipe for trouble," Gold explains. He writes that during the stock-market crash of 2008, Chesapeake's shares lost 59% of their value; subsequently, there were revelations that McClendon had taken out large personal loans from EIG Global Energy Partners (an investor Eads had connected with Chesapeake) without the knowledge of his company's board. Large shareholders became increasingly restive, he writes. McClendon's job was terminated in 2013.
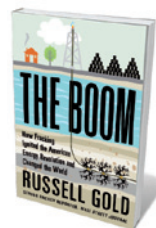
The Boom: How Fracking Ignited the American Energy Revolution and Changed the World
RUSSELL GOLD
Simon & Schuster: 2014.

The second story that Gold details is of a secretive partnership between McClendon and Carl Pope, head of the Sierra Club, the United States' largest environmental group, for 18 years. The two had a mutual interest in fighting coal power, but very different motivations: Pope's, to combat climate change; McClendon's, to secure natural gas as the nation's power-grid fuel of choice. Without the knowledge of the Sierra Club board, Gold alleges, Pope allowed McClendon to underwrite the organization's Beyond Coal campaign, even as its New York chapter was fighting fracking over water contamination. Gold describes how tension within the Sierra Club grew. Pope was ousted in 2011.

That the costs and benefits of fracking are complex is never fudged. The rapid decline rates of wells for shale gas and 'tight', or fracked, oil condemn frackers to increase the pace of drilling continually just to keep overall production flat. Embedded in the unfettered, market-driven US approach to energy development is the risk of dependence on finite resources. The US Energy Information Administration forecasts that tight oil will peak by 2021, although some think this optimistic.

The Boom carefully explores technical issues around water contamination and the failure to develop solutions. The inability of regulators to keep up with entrepreneurs' advances — a problem as old as the oil industry itself — opens up scenarios of irreversible damage. And drilling has "thrown a lifeline to fossil fuels" just when dependence on them, and the risks of climate change, are more dangerous than ever.

Gold points out that over the 150-year history of oil and gas exploitation, we have worked our way down from vast accumulations of cheap, high-quality hydrocarbons to poor-quality shales. "Source rock is where plankton turned into hydrocarbons," he muses. "There is no further back." We "fossil-fuel addicts", he suggests, should think of shale gas as methadone, a stepping stone on the way to a renewably powered future. ∎

Chris Nelder *is an independent energy analyst and journalist. He wrote* Profit from the Peak *and* Investing in Renewable Energy, *and he blogs at GetREALList.com.*
e-mail: chris@getreallist.com

The Serpent's Promise: The Bible Retold as Science
Steve Jones (Little, Brown, 2014)
With sensitivity towards religion and sardonic wit, geneticist Steve Jones delivers a masterful scientific take on biblical events such as the Deluge — which he attributes to the end of an ice age. (See Tim Radford's review: *Nature* **496,** 432–433; 2013.)
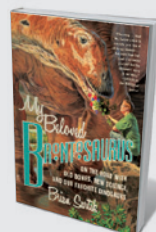
My Beloved Brontosaurus
Brian Switek (Farrar, Straus and Giroux, 2014)
The film *Jurassic Park* aside, velociraptors were turkey-sized, reveals Brian Switek in this paean to palaeontology. He joins the scales-versus-feathers debate, and mourns the 'second extinction' of the brontosaurus, now called apatosaurus. (See Xu Xing's review: *Nature* **496,** 30; 2013.) **Emily Banham**

Correspondence

Growing time lag threatens Nobels

The time lag between reporting a scientific discovery worthy of a Nobel prize and the awarding of the medal has increased, with waits of more than 20 years becoming common. If this trend continues, some candidates might not live long enough to attend their Nobel ceremonies.

Before 1940, Nobels were awarded more than 20 years after the original discovery for only about 11% of physics, 15% of chemistry and 24% of physiology or medicine prizes, respectively. Since 1985, however, such lengthy delays have featured in 60%, 52% and 45% of these awards, respectively.

The increasing average interval between reporting discoveries and their formal recognition can be fitted to an exponential curve (see 'The long road to Sweden'), with data points scattered about the mean value.

As this average interval becomes longer, so the average age at which laureates are awarded the prize goes up. By the end of this century, the prizewinners' predicted average age for receiving the award is likely to exceed his or her projected life expectancy (data not shown). Given that the Nobel prize cannot be awarded posthumously, this lag threatens to undermine science's most venerable institution.

Santo Fortunato* *Aalto University, Finland.*
santo.fortunato@gmail.com
On behalf of 6 co-authors; see go.nature.com/cmmxa5 for full list.

Livestock: tackle demand and yields

Among many otherwise laudable suggestions, Mark Eisler and colleagues propose limiting feedstuffs for livestock to fibrous fodder, such as grass and silage (see *Nature* **507,** 32–34; 2014). However, we believe that any attempt to meet the rapid growth

THE LONG ROAD TO SWEDEN

Scientists who publish prizewinning discoveries are, on average, waiting longer for a Nobel than ever before.

Data points are 5-year-averaged waits; shading around lines shows confidence limits.
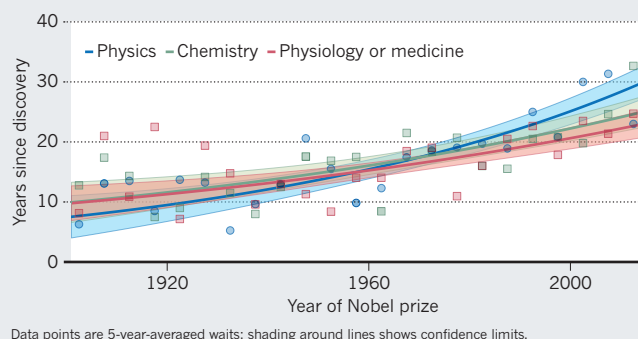
in world demand for meat and dairy products by focusing on ruminant grazing systems would be damaging for biodiversity and for the global climate.

Although ruminants convert grass and silage into animal protein, they do so inefficiently; they therefore require much more land to produce a given amount of meat or milk than ruminants fed on diets that include grain. Growing enough fodder to satisfy demand would require the large-scale expansion of grazing lands (see go.nature.com/7mf63y) — a leading cause of biodiversity loss, tropical deforestation and carbon dioxide emissions.

The environmental impacts of meat and dairy production should instead be addressed by stringent efforts to decrease consumption, halt the expansion of grazing, and increase yields on land that is already used for livestock. Promoting extensive grazing without tackling demand would do more harm than good.

Erasmus K. H. J. zu Ermgassen, David R. Williams, Andrew Balmford *University of Cambridge, UK.*
ekhjz2@cam.ac.uk

Livestock: limit red meat consumption

Mark Eisler and co-authors advocate eating only 300 grams of red meat a week (roughly the volume of three decks of playing cards) as a step towards producing sustainable livestock (*Nature* **507,** 32–34; 2014). That amount corresponds to 3.5–7% of a 2,000-calorie-a-day diet, depending on the cut and type of meat. Such a move would also make for a more equitable global distribution of animal-product consumption; these products comprise around 48% of the average diet in the United States, for example (S. Bonhommeau *et al. Proc. Natl Acad. Sci. USA* **110,** 20617–20620; 2013).

Imposing a global dietary limit of 5% red meat as part of a 10% maximum for all animal-based products would enable more people to be fed using less land. For example, eliminating livestock and using existing agricultural lands to grow crops for direct human consumption instead of for livestock fodder could feed an extra 4 billion people (E. S. Cassidy *et al. Environ. Res. Lett.* **8,** 034015; 2013), thereby reducing or eliminating the greenhouse-gas emissions and biodiversity loss associated with conversion of natural habitats. This would also reduce many other environmental impacts of agriculture that relate to the use of water, fertilizer and fossil fuels.

Brian Machovina, Kenneth J. Feeley *Florida International University, Miami; and The Fairchild Tropical Botanic Garden, Coral Gables, Florida, USA.*
brianmachovina@gmail.com

Zoo visits boost biodiversity literacy

Zoos and aquaria worldwide attract more than 700 million visits every year. They are therefore well placed to make more people aware of the importance of biodiversity — a prime target of the United Nations Strategic Plan for Biodiversity 2011–20.

We surveyed approximately 6,000 visitors to 30 zoos and aquaria in 19 countries (see go.nature.com/vwf8yf). More respondents showed improved understanding of biodiversity after their visit (75.1% compared with 69.8% before) and more could identify an individual action that would bolster biodiversity after their visit (58.8% compared with 50.5% before).

Regrettably, increased awareness does not necessarily change behaviour. The world's zoo and aquarium communities must also help to drive important behavioural and social changes to assist conservation.

Andrew Moss *Chester Zoo, UK.*
Eric Jensen *University of Warwick, Coventry, UK.*
Markus Gusset *World Association of Zoos and Aquariums, Gland, Switzerland.*
markus.gusset@waza.org

A protein that spells trouble

The gene *CYLD* is so named because one of its mutant forms is associated with cylindromatosis, which causes skin tumours.

The CYLD protein is an enzyme; its active site in humans contains a cysteine residue at position 601 (denoted as C in the one-letter amino-acid code). The amino-acid sequence following this cysteine (C) is tyrosine (Y), leucine (L) and aspartate (D). What are the odds of that?

David Boone *Indiana University School of Medicine — South Bend, Indiana, USA.*
daboone@iu.edu

Alejandro Zaffaroni

(1923–2014)

Bioentrepreneur who revolutionized drug delivery and screening.

Alejandro Zaffaroni spent some 60 years helping to conduct research that led to the birth-control pill, and co-founded nearly a dozen biotechnology start-ups in California's Silicon Valley. He commercialized innovative ways to get drugs into the body. His adhesive skin patches, implanted devices and controlled-release capsules reduced side effects, boosted efficacy and delivered the correct dose of medicine without complicated schedules.

Zaffaroni, who died on 1 March aged 91, was born in Montevideo, Uruguay. He studied for his bachelor's degree there, and in the last days of the Second World War he boarded a military cargo ship to New York, where he gained a PhD in biochemistry at the University of Rochester. He conducted postdoctoral research under a fellowship from the US National Institutes of Health. Zaffaroni adapted a technique known as paper chromatography to purify steroids — the group of chemicals that includes cholesterol and sex hormones. The research led to the first synthesis of the steroid hormone cortisone.

Continuing his work on steroid biosynthesis, Zaffaroni joined Syntex, a small Mexican firm extracting steroid precursors from yams to make therapeutic hormones. He helped to grow Syntex into a pharmaceutical company, and opened a US subsidiary in Palo Alto, California, in 1962. Studying steroid-containing ointments to treat skin disorders, Zaffaroni realized that significant amounts of steroids entered the bloodstream through the skin. He founded the ALZA Corporation in 1968 to control and capitalize on this effect.

Zaffaroni pursued risky businesses and persevered through pitfalls. When a patch for motion sickness was developed at ALZA, the head of marketing protested that it would be difficult to commercialize. But Zaffaroni knew that it would be a springboard; he predicted that the product would gain approval from the US Food and Drug Administration, opening doors to collaborations with deeper-pocketed pharmaceutical partners.

ALZA went on to develop skin patches to help people to quit smoking and to treat pain and high blood pressure, among other conditions. It developed an intrauterine contraceptive; a polymer film that delivered glaucoma drugs without blurring vision; and swallowable capsules that released drugs over many hours. Pharmaceutical giant Johnson & Johnson bought ALZA for US$12.3 billion in 2001.

The foresight that prompted Zaffaroni to

ROGER RESSMEYER/CORBIS
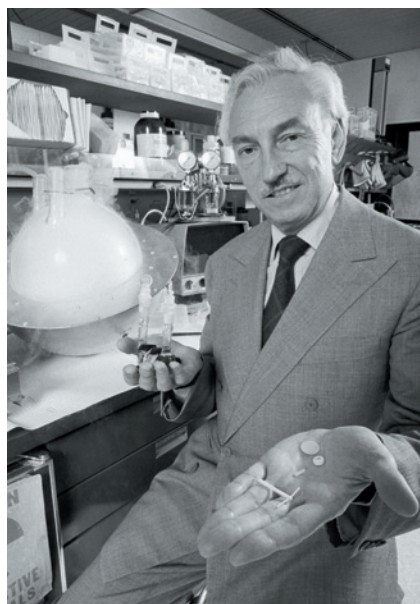
leave Syntex and start ALZA led to a model for his other companies. First he would articulate a challenge in conversations with industrial and academic scientists. Then he would systematically evaluate existing technologies, competition and potential human impact. When ready to launch a company, he would seek the best advisers and scientific talent to translate basic discoveries into applications.

Zaffaroni did not supervise his scientific teams closely. He empowered people to achieve great things. He paid attention to building design to ensure cross-disciplinary interactions, and he considered colours of furnishings and artwork — all of which he believed would stimulate creativity. More than 40 ALZA employees went on to serve as chief executives of other companies.

The youngest of five children, Zaffaroni credited his mother, a teacher, for his creativity and ability to combine different disciplines. To his father, a banker, he attributed his business acumen and his ability to select, trust and galvanize good people.

The people he tapped would accompany him in pursuits that they had never imagined. In 1970, a phone call from Zaffaroni brought me to Palo Alto from my laboratory at what was then the Worcester Foundation for Experimental Biology in Shrewsbury, Massachusetts. I knew nothing of the pharmaceutical industry and was not sure what job he was offering. But his passion and commitment inspired me to leave my career in

research for one in pharmaceuticals. Like others, I thought, "If Dr Zaffaroni is leading, I will follow." I did, with not a single regret.

In 1980, Zaffaroni established the DNAX Research Institute of Molecular and Cellular Biology in Palo Alto with three distinguished scientists (two of them Nobel laureates) from Stanford University in California to apply cellular and molecular biology to medicine. This immunology-research paradise fostered the discovery of several signalling proteins used by white blood cells. DNAX was acquired by drug company Schering-Plough in 1982 for $29 million — reportedly $1 million for each employee with a PhD.

In his mid-60s, Zaffaroni turned his attention to making drug discovery more efficient, and founded Affymax in 1988. The firm borrowed technology for making computer chips to produce microarrays that rapidly make and test many potential drugs against their protein targets. This was probably the first industrialized biochemical research, and the technique — combinatorial chemistry — is now a component in many major drug-discovery programmes.

The technology behind the chips at Affymax led Zaffaroni to spin out a new company, Affymetrix, in 1991. This was the first company to develop DNA microarrays, which allow rapid analysis of mutations and gene activity. These enabled the rise of genome-wide association studies and are now used widely to detect genetic variants. Subsequent discoveries yielded a string of companies, including Verdia, Avidia, Codexis and Symyx. Maxygen, for example, worked on 'shuffling' DNA to iteratively improve protein drugs.

The last company that Zaffaroni founded was Alexza in 2000. Its work was inspired by what he identified as the most efficient of all drug-delivery systems: the cigarette. Noting the rapid effect of inhaled nicotine, he sought to apply it to other drugs. Alexza's first inhaled product, for treatment of acute agitation in adults with schizophrenia or bipolar disorder, was approved last year.

As he shaped the industry, Zaffaroni's companies grew, but he remained a famously understated entrepreneur. His impact on biotechnology is immeasurable. ∎

Jane E. Shaw *is retired chairman of the board of Intel. She was an employee of ALZA from 1970 to 1994 and served as president and chief operating officer from 1989 to 1994.
e-mail: janeeshaw@comcast.net*

NEWS & VIEWS

Cracking up on asteroids

A combination of laboratory experiments and modelling shows that diurnal temperature variations are the main cause of rock breakdown and the ensuing formation of powdery rubble on the surface of small asteroids. See Letter p.233

HEATHER A. VILES

The surfaces of asteroids orbiting the Sun are known to be covered in particulate debris called regolith, rather like the non-living components of soil on Earth. Impacts by meteoroids — meteorites when they reach the ground — and micrometeoroids are usually thought to be the major agents of regolith formation on rocky bodies that have no atmosphere, such as asteroids. Meteoroid impacts rupture boulders, producing smaller debris, whereas micrometeoroids (up to a few millimetres in diameter) cause the gradual erosion of boulders and debris into finer particles. On page 233 of this issue, Delbo *et al.*[1] use data from laboratory experiments and modelling to test the alternative hypothesis that temperature cycling is the major cause of regolith production on small (kilometre-sized and smaller) asteroids and may contribute to it on larger asteroids (Fig. 1).

The ability of diurnal temperature cycling to cause the mechanical breakdown of surface rocks and boulders on Earth and other planetary bodies has been heavily debated for more than a century. In essence, directional heating and cooling by solar radiation can cause large temperature changes at the surface of rocks and boulders. Depending on the thermal conductivity of the materials involved, large temperature gradients can occur between the surface and inner portions of the boulders, with resultant mechanical stresses that, when repeated over time, can lead to cracking and deterioration of the boulders.

Studies[2–4] based on field observations, laboratory experiments and modelling have confirmed the effectiveness of such thermal weathering in a range of settings (Earth's deserts, on Mars and on other planetary bodies), but have not evaluated whether it is the dominant process. Delbo *et al.* have made a major advance by linking results from experimental weathering of meteorites to a micromechanical model. This allowed them to quantify the rate of breakdown from crack growth caused by thermal cycling and to compare it with the rate of attrition by micrometeoroid impact.

The authors performed experiments on samples of two chondrites (stony meteorites)

Figure 1 | Asteroid Ida. This enhanced composite image of Ida, a silicate asteroid more than 50 kilometres in length within the main asteroid belt, was taken by the camera aboard the Galileo spacecraft on 28 August 1993. Colour differences in the image reflect both the steep topography and the varied physical state and composition of the regolith on Ida. Delbo *et al.*[1] show that temperature cycling is the main source of regolith formation on kilometre-sized and smaller asteroids.

NASA/JPL

approximately 1 centimetre in diameter — one an ordinary (silicate) chondrite and the other a carbonaceous chondrite, which is dark and has low reflectivity. They exposed them to more than 400 diurnal temperature cycles representative of conditions found on near-Earth asteroids, each cycle having a temperature range of 190 kelvins and lasting 2.2 hours. Using the technique of X-ray computed tomography scanning to calculate rates of crack growth in the samples, Delbo *et al.* observed a measurable increase in the size of pre-existing cracks in both types of meteorite even after 76 temperature cycles. Their micromechanical model simulates the measured growth of cracks and shows how growth from an initial crack 30 micrometres in length leads to rapid breakdown of centimetre-sized boulders and the production of finer-grained particulate material.

For silicate asteroids at 1 astronomical unit (AU), where 1 AU is the mean Earth–Sun distance, the authors' model predicts breakdown rates that are at least tenfold quicker than those caused by micrometeoroid impact, with larger boulders predicted to break down faster than

smaller ones. The disparity is even larger for the carbonaceous asteroids at 1 AU: a 10-cm boulder is predicted to survive less than 1,000 years under thermal cracking, compared with 10 million years under micrometeoroid attrition. These findings may explain why carbonaceous asteroids in near-Earth orbits are rare. Even for boulders on asteroids farther from the Sun, for example in the inner parts of the main asteroid belt, which lies between the orbits of Mars and Jupiter, the model's results show that thermal cycling causes faster breakdown than micrometeoroid impacts.

Although Delbo and colleagues' findings throw interesting light on the role of diurnal heating and cooling in the breakdown of rocks on planetary bodies, they require further testing and development. Experiments on meteorites under controlled laboratory conditions are necessarily imperfect analogues of conditions and materials on asteroid surfaces. It would be useful to have a wider range of experimental results with which to compare model predictions. Longer experiments than those performed by the authors would be particularly helpful, to try to capture the shift from crack

development to boulder breakdown, which is predicted by their model to occur after between 10^7 and 10^8 experimental cycles. Furthermore, it would be good to develop more-sophisticated modelling to include situations in which multiple cracks develop in single boulders and produce more rapid breakdown. It might also be fruitful to investigate potential synergies between micrometeoroid impact and thermal-weathering processes.

Delbo and co-workers' insightful work should lead to further advances in understanding rock breakdown and regolith formation on asteroids and other rocky bodies that have no atmosphere. Their work has obvious ramifications for developing the study of the geomorphology of asteroid surfaces. Diurnal temperature cycling has previously been neglected as a potential contributor to surface modification and regolith production on asteroids. It should now be considered alongside other processes of space weathering that can produce debris of various sizes and characteristics, and thus affect the interpretation of remotely sensed spectroscopic data from asteroids[5–7].

The work is also relevant for the investigation of rock breakdown and regolith formation on Earth and Mars. Although conditions are more complex on planetary surfaces with an atmosphere, the methodology developed by Delbo et al., which links laboratory experimental quantification of crack growth to micromechanical modelling, could be used to good effect to explore the role of diurnal temperature cycling in the breakdown of surface boulders under arid conditions. ■

Heather A. Viles is in the School of Geography and the Environment, University of Oxford, Oxford OX1 3QY, UK.
e-mail: heather.viles@ouce.ox.ac.uk

1. Delbo, M. et al. Nature 508, 233–236 (2014).
2. Viles, H. A. et al. Geophys. Res. Lett. 37, L18201 (2010).
3. Molaro, J. & Byrne, S. J. Geophys. Res. 117, E10011 (2012).
4. Eppes, M. C., McFadden, L. D., Wegmann, K. W. & Scuderi, L. A. Geomorphology 123, 97–108 (2010).
5. Clark, B. E., Hapke, B., Pieters, C. & Britt, D. in Asteroids III (eds Bottke, W. F. Jr, Cellino, A., Paolicchi, P. & Binzel, R. P.) 585–599 (Univ. Arizona Press, 2002).
6. Pieters, C. M. et al. Nature 491, 79–82 (2012).
7. Gaffey, M. J. Icarus 209, 564–574 (2010).

This article was published online on 2 April 2014.

CANCER

Damage prevention targeted

The MTH1 protein prevents oxidized nucleotides from being misincorporated into DNA. Two studies find that selective inhibition of MTH1 by small molecules suppresses tumour growth. SEE ARTICLES P.215 & P.222

DAN DOMINISSINI & CHUAN HE

Cellular metabolism produces reactive oxygen species that can cause oxidative damage to DNA, RNA, proteins and lipids. The metabolism of cancer cells is faster than that of normal cells, and they therefore need to cope with higher levels of these molecules[1]. This not only threatens nucleic acids, but also endangers the cellular pool of nucleoside triphosphates (NTPs; nucleotides that contain three phosphate groups), potentially producing oxidized DNA bases that could be misincorporated into DNA and RNA[2]. The MTH1 enzyme alleviates this problem by selectively removing oxidized nucleotides from the NTP pool[3]. Two papers[4,5] in this issue show that MTH1 is not essential for normal cells, as was previously suggested[6], but is required for cancer-cell survival. The authors also report that selective MTH1 inhibition by small molecules leads to DNA damage and effectively suppresses cancer growth, thus revealing this protein as a promising target for anticancer therapies.

The free NTPs used in DNA synthesis (dNTPs) are up to 13,000-fold more susceptible to oxidative damage than are bases in duplex DNA[7]. If oxidized dNTPs were incorporated into DNA, the resulting nucleic acid would carry hundreds of times more damaged bases than would result from oxidation of the whole DNA molecule. This could have severe consequences, such as genome instability and mutations. Almost all organisms have evolved pathways to counteract the potentially deleterious effects of misincorporation of oxidized nucleotides into nucleic-acid polymers. In humans, the MTH1 protein hydrolyses 8-oxo-dGTP and 2-OH-dATP (oxidized forms of the nucleotides dGTP and dATP, respectively), converting them into monophosphate products that cannot be used for DNA synthesis[8] (Fig. 1).

Cancer cells maintain an altered redox state[9] with a typically higher overall burden of oxidative stress than is found in normal cells, which could be key to the cause and progression of cancer[10]. So there is a compelling case for designing an anticancer strategy that exploits the redox difference between normal and cancer cells by selectively disabling the mechanisms that allow cancer cells to adapt to their redox state.

Gad et al.[4] (page 215) have made great strides in this direction. They speculated that MTH1 might have a crucial redox-adaptive role in cancer cells on the basis of two previous observations: that overexpression of MTH1 prevents DNA damage induced by the cancer-causing RAS protein, as well as the attendant premature cellular senescence[11]; and that such overexpression also lowers the high mutation rate and genome instability typical of colorectal cancer cells that have a defective mechanism for repairing base mismatches in DNA[12]. The authors therefore tested the hypothesis that the normally non-essential
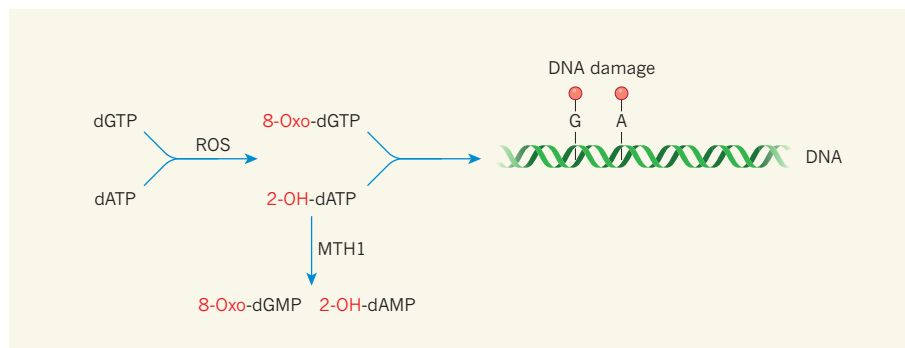
Figure 1 | Role of the MTH1 protein. In cells, reactive oxygen species (ROS) convert the free nucleotides dGTP and dATP to the oxidized triphosphates 8-oxo-dGTP and 2-OH-dATP, respectively. If incorporated into DNA, the oxidized G and A bases of 8-oxo-dGTP and 2-OH-dATP represent forms of DNA damage. MTH1 hydrolyses the oxidized triphosphates into corresponding monophosphates (8-oxo-dGMP and 2-OH-dAMP), which cannot be incorporated into DNA. Gad et al.[4] and Huber et al.[5] find that MTH1 is non-essential in normal cells, but that in cancer cells, which have higher levels of ROS, its inhibition results in the incorporation of oxidized nucleotides into DNA, causing DNA damage and cell death.

MTH1 protein could be required for the survival of cancer cells because it protects the dNTP pool from oxidative damage.

Using a combination of experimental approaches *in vitro* and *in vivo*, the researchers confirmed that hypothesis. The authors also developed two small molecules that potently and selectivity inhibit MTH1 *in vivo* by binding to its active site. As expected, the inhibitors cause increased incorporation of oxidized dNTPs, DNA damage, cytotoxicity and a therapeutic response in models of cancer, but not in normal cells.

In a parallel study, Huber *et al.*[5] (page 222) came up with the same anticancer target, but by a different path. The small molecule SCH51344 has previously been reported to suppress cancers induced by RAS proteins[13]. The authors identified MTH1 as the molecule's target by incubating an analogue of SCH51344 as 'bait' with the contents of cancer cells known to be sensitive to the compound. Because SCH51344 has not been tested in clinical trials, they extended their screen to search for other MTH1 inhibitors that are likely to have good pharmacological properties *in vivo*. They identified crizotinib — a drug that was approved in 2011 for a type of lung cancer[14] — as a potent inhibitor of MTH1. Surprisingly, it was not the clinically used isomer (enantiomer) of crizotinib that was active against MTH1, but the mirror-image version.

Huber *et al.* went on to manipulate MTH1 activity in cancer models using their inhibitors, and came to similar conclusions to Gad and co-workers. Taken together, the two papers importantly show that neither the activity of pivotal genome 'guardians' (such as the p53, ATM and ATR proteins) nor the expression level of RAS proteins seem to be major determinants of the models' response to the MTH1 inhibitors, suggesting that the compounds will be therapeutically useful against tumours in which these guardians are compromised.

Both studies present compelling evidence that MTH1 inhibition suppresses cancer growth by preventing hydrolysis of 2-OH-dATP and 8-oxo-dGTP. However, human MTH1 has higher hydrolytic activity for 2-OH-dATP and 2-OH-rATP (an oxidized form of the RNA nucleotide rATP) than for 8-oxo-dGTP, and its affinity for 2-OH-rATP is the highest among all known substrates[15]. Considering that the cellular rATP pool is about 30-fold larger than that of dATP and roughly 350-fold larger than that of dGTP (ref. 16), and that treatment of rATP with reactive oxygen species generates 2-OH-rATP effectively[15], the latter could be a major cellular substrate of human MTH1. Misincorporation of 2-OH-rATP into RNA molecules could also have deleterious consequences[17]. Such consequences have received little attention, presumably because of the idea that the transient nature and rapid turnover of RNA molecules render the effects of their damage negligible. However, mounting evidence from work *in vitro*

and *in vivo*, especially in neurodegenerative diseases, suggests otherwise[18].

More broadly, rATP and rGTP constitute the largest and third-largest pools, respectively, of any NTP in mammalian cells[16], and, besides their role in RNA synthesis, have functions in energy metabolism and signal transduction. Oxidized rATP or rGTP might therefore interfere with regulatory signalling or other cellular functions performed by undamaged triphosphates. Such effects should not be overlooked, and their potential contribution to the suppression of cancer-cell growth through MTH1 inhibition should be evaluated in the future.

With renewed interest in tumour metabolism, altered redox states have emerged as attractive targets that extend beyond gene- or pathway-specific mutations as a means of tackling the unique but widely encountered biochemical anomalies in cancer cells[19,20]. Further studies of these targets could thereby address problems inherent to personalized approaches for anticancer therapy — such as the existence of different types of cancer cell within individual tumours and the development of drug resistance associated with rapid evolution of cancer cells. By aiming at a redox-adaptation mechanism, MTH1 inhibition embodies a new approach for this strategy, one that is poised to directly affect both DNA and RNA synthesis, and potentially also bioenergetic and signalling pathways. ∎

Dan Dominissini *and* **Chuan He** *are in the Department of Chemistry and Institute for Biophysical Dynamics, and at the Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois 60637, USA.*
e-mails: dominissinid@uchicago.edu; chuanhe@uchicago.edu

1. Cairns, R. A., Harris, I. S. & Mak, T. W. *Nature Rev. Cancer* **11**, 85–95 (2011).
2. Luo, M., He, H., Kelley, M. R. & Georgiadis, M. M. *Antioxid. Redox Signal.* **12**, 1247–1269 (2010).
3. Yoshimura, D. et al. *J. Biol. Chem.* **278**, 37965–37973 (2003).
4. Gad, H. et al. *Nature* **508**, 215–221 (2014).
5. Huber, K. V. M. et al. *Nature* **508**, 222–226 (2014).
6. Tsuzuki, T., Egashira, A. & Kura, S. *Mutat. Res.* **477**, 71–78 (2001).
7. Topal, M. D. & Baker, M. S. *Proc. Natl Acad. Sci. USA* **79**, 2211–2215 (1982).
8. Fujikawa, K. et al. *J. Biol. Chem.* **274**, 18201–18205 (1999).
9. Toyokuni, S., Okamoto, K., Yodoi, J. & Hiai, H. *FEBS Lett.* **358**, 1–3 (1995).
10. Zhang, Y. et al. *Antioxid. Redox Signal.* **15**, 2867–2908 (2011).
11. Rai, P. et al. *Oncogene* **30**, 1489–1496 (2011).
12. Russo, M. T. et al. *Mol. Cell. Biol.* **24**, 465–474 (2004).
13. Kumar, C. C. et al. *Cancer Res.* **55**, 5106–5117 (1995).
14. Camidge, D. R. et al. *Lancet Oncol.* **13**, 1011–1019 (2012).
15. Fujikawa, K., Kamiya, H., Yakushiji, H., Nakabeppu, Y. & Kasai, H. *Nucleic Acids Res.* **29**, 449–454 (2001).
16. Bestwick, R. K., Moffett, G. L. & Mathews, C. K. *J. Biol. Chem.* **257**, 9300–9304 (1982).
17. Tanaka, M., Chock, P. B. & Stadtman, E. R. *Proc. Natl Acad. Sci. USA* **104**, 66–71 (2007).
18. Nunomura, A. et al. *Neurotox. Res.* **22**, 231–248 (2012).
19. Trachootham, D., Alexandre, J. & Huang, P. *Nature Rev. Drug Discov.* **8**, 579–591 (2009).
20. Raj, L. et al. *Nature* **475**, 231–234 (2011).

This article was published online on 2 April 2014.

Africa's greenhouse-gas budget is in the red

Africa had been thought to be a potentially large carbon sink — of great value in efforts to mitigate carbon dioxide emissions. But an analysis now reveals that it could be a net source of greenhouse gases that will increase global warming.

CHEIKH MBOW

One of the biggest challenges in curbing climate change is to obtain robust estimates of greenhouse-gas emissions and sequestration. Such estimates are necessary for predicting the effects of land-use-related mitigation efforts on the release and uptake of the three main greenhouse gases — carbon dioxide, nitrous oxide and methane — produced by complex terrestrial processes. Writing in *Biogeosciences*, Valentini *et al.*[1] have risen to this challenge by providing a full greenhouse-gas assessment for Africa.

Their paper addresses the burning need for evidence-based knowledge[2] to devise effective policies for reducing the impact of human activities on climate and the effects of climate change on societies, and to decide where investment is required to meet those aims.

Few reliable estimates of carbon flux are available for Africa, because information about its carbon budget is partial, fragmented and inconsistent. Furthermore, greenhouse-gas budgets generally vary in precision and clarity because of the different methods used to produce them. Existing data are therefore subject to many uncertainties associated with

incomplete data sets, short observation periods and unknown fluxes from land-use change and forest fires. Other problems include the limited data available to calibrate the results of modelling studies, and the empirical assumptions made about factors such as deforestation rates, carbon density in major pools and losses of carbon from soil.

Valentini and colleagues used several approaches to solve these problems. They include: data collection from ground-level observations, such as studies of biomass, tree measurements and ecosystem inventories of gas fluxes (Fig. 1); satellite-based data gathering; modelling of carbon fluxes; and derivation of emissions data from secondary sources of information, such as records of agricultural activity. In this way, the authors show that Africa is central to global mitigation efforts because of its large forested areas, the diversity of its ecosystems and the volume of emissions[3] it produces as a result of land use and land-use change.

Until now, scientific opinion has held that Africa could help to reduce emissions or sequester carbon if deforestation of large areas, such as the Congo Basin, could be avoided, or if tree and forest cover could be increased through sustainable practices, for example agroforestry and plantation management. The authors concur that Africa is a small carbon sink on an annual timescale. But, more surprisingly, they find that it may be a net source of radiative forcing — reradiation of heat back towards Earth's surface by greenhouse-gas molecules — when methane and nitrous oxide are included in the annual budgeting.

One of the merits of this study is that Valentini et al. disaggregate their data, determining the flux of greenhouse gases from different carbon pools (biomass and soil carbon) and land-cover classes (such as croplands, forests and shrub lands), and from the sources and sinks associated with various forms of land use. Furthermore, they establish the temporal evolution of the net budget for different forms of land use in the four sub-Saharan African regions. They find that the emission rate of carbon dioxide from the continent was nearly constant between 1980 and 2009, but that there were regional variations: western Africa had the largest increase of emissions and eastern Africa the lowest; only central Africa had decreased emissions during that period. These differences might reflect the regional frequency and magnitude of vegetation fires, which are used mostly to expand cropland and shift areas of cultivation. The use of firewood, or of fire in land management, leads to substantial carbon emissions because biomass is rapidly oxidized.

Are these conclusions merely implications associated with data uncertainty, or plausible outcomes of an integrated approach? There is

Figure 1 | Flux tower at Dahra, Senegal. Valentini et al.[1] have used data from several sources, including information about gas fluxes obtained from facilities such as this one, to determine a full greenhouse-gas budget for Africa.

previous evidence to suggest the latter, such as the finding[4] that emissions produced by tropical deforestation are only partially offset by the carbon sequestration associated with forest regrowth.

Valentini and colleagues' results are crucial to the debate about greenhouse-gas emissions related to land use in Africa[2], where most 'game changers' are occurring through shifts in either agricultural development models (family farming to agri-business; labour intensive to capital intensive; small scale to large scale) or production systems (farming and agriculture versus conservation)[5]. These structural changes in agricultural systems were not considered explicitly in the paper, but will need more attention if we are to work out precisely how each one affects greenhouse-gas emissions. If emissions tend to increase because of these changes, then strategies could be developed to address either the specific land-use processes that cause emissions, or, most importantly, the drivers of change. This approach is central to climate-smart agriculture, an emerging concept that minimizes greenhouse-gas emissions by encouraging sustainable farming and land-use practices.

This paper could also instigate a turning point in the design of programmes for counteracting climate change in Africa. Such programmes are currently based mainly on the assumption[6,7] that agriculture, forestry and other land uses (AFOLU) collectively form the main terrestrial carbon sink. In Africa,

however, emissions from land-use change are exceptionally high, and are much greater than those from burning fossil fuels. This makes Africa unique among the continents. However, devising new programmes on the basis of Valentini and co-workers' findings will be difficult, because accounting for the three major greenhouse gases in the AFOLU sector requires high investment and skills not always available in developing countries[2,8].

Valentini et al. did not fully consider two main aspects of the African carbon cycle. The first is the key role of agroforestry, which could potentially sequester large amounts of carbon, reduce deforestation and provide many ecosystem services, such as wood fuel[9]. The second is the part played by peat lands and by vegetation in flood plains. Although such regions are small in terms of surface area, they produce substantial carbon emissions, especially when they undergo land-use changes. This is particularly true when pristine peat lands are converted[10] to fields for agriculture or plantations. Further studies are needed to determine whether either of these aspects of the carbon cycle could change the outcome of the authors' findings.

Other future directions of research should include the creation of an inventory of databases of information from in situ studies, as part of a collaborative framework to improve the availability of low-cost greenhouse-gas data. Such frameworks, which are being promoted worldwide, help to harmonize data-collection methods and improve the consistency of the collected data. The advent of new generations of satellites also provides opportunities to substantially improve carbon budgeting for Africa, and possibly the budgeting of other greenhouse gases. ■

Cheikh Mbow *is at the World Agroforestry Centre (ICRAF), PO Box 30677, Nairobi 00100, Kenya.*
e-mail: c.mbow@cgiar.org

1. Valentini, R. *et al. Biogeosciences* **11,** 381–407 (2014).
2. Olander, L., Wollenberg, E., Tubiello, F. & Herold, M. *Environ. Res. Lett.* **8,** 011002 (2013).
3. Brown, S., Grais, A., Ambagis, S. & Pearson, T. in *Baseline GHG Emissions from the Agricultural Sector and Mitigation Potential in Countries of East and West Africa* CCAFS Working Paper 13 (CCAFS, 2012).
4. Pan, Y. *et al. Science* **333,** 988–993 (2011).
5. Haberl, H. *et al.* in *Rethinking Global Land Use in an Urban Era* (eds Seto, K. C. & Reenberg, A.) Ch. 4, 35–69 (MIT Press, 2014).
6. Zomer, R. J., Trabucco, A., Bossio, D. A. & Verchot, L. V. *Agric. Ecosyst. Environ.* **126,** 67–80 (2008).
7. Saatchi, S. S. *et al. Proc. Natl Acad. Sci. USA* **108,** 9899–9904 (2011).
8. Franks, J. R. & Hadingham, B. *Land Use Policy* **29,** 727–736 (2012).
9. Mbow, C., Smith, P., Skole, D., Duguma, L. & Bustamante, M. *Curr. Opin. Environ. Sustainability* **6,** 8–14 (2014).
10. Danielsen, F. *et al. Conserv. Biol.* **23,** 348–358 (2009).

METABOLISM

Targeting a fat–accumulation gene

An enzyme that links two metabolic hubs has been found to be upregulated in the fat cells of overweight mice. Inhibition of the gene encoding this enzyme protects mice from diet–induced obesity. See Letter p.258

CHARLES BRENNER

In recent decades, increased food availability has dramatically altered many people's diets. Each year, billions of tonnes of grain are fed to livestock or converted to ingredients for food. Consumption of the resulting meat, fat and processed carbohydrate by humans, accompanied by a decline in their energy expenditure, has created a nutritional overload that has resulted in increased adiposity and incidence of the resulting diseases worldwide. On page 258 of this issue, Kraus et al.[1] identify *Nnmt,* which encodes the enzyme nicotinamide *N*-methyltransferase (NNMT), as a gene whose activity is ramped up by nutritional overload, and they find that *Nnmt* is required for the accumulation of fat.

Fat storage is essential for animal development and survival, particularly in times of macronutrient (protein, fat and carbohydrate) limitation, which almost certainly drove 500 million years of animal evolution[2]. The insulin system allows animals to convert excess carbohydrate to fat — the most energy-rich component of an animal's body mass — in part by instructing adipose tissue to take up glucose from the bloodstream.

The ability to increase body mass on a high-fat diet is a trait that has increased animal survival throughout evolution. Only in circumstances of chronic over-nutrition does our ability to convert sugar to fat and to store excess fat predispose us to disease rather than to resistance to famine.

When the gene encoding the insulin-sensitive glucose-transporter protein GLUT4 is deleted from a mouse's fat cells, the animal develops insulin resistance in several tissues and cannot effectively clear blood glucose[3]. Because insulin stimulates GLUT4 activity and because GLUT4 is also required for insulin sensitivity, alteration of the GLUT4–insulin circuit has powerful effects on metabolism and can lead to diabetes. Components of this circuit could therefore emerge as drug targets to mitigate the effects of over-nutrition on human health.

Mice that lack or overexpress GLUT4 in fat are used as models for people sensitized to or resistant to diabetes, respectively[4]. To identify molecules involved in the development of diabetes, Kraus and colleagues used these animals to find genes that are highly expressed when GLUT4 is deleted and that are repressed when GLUT4 is overexpressed.
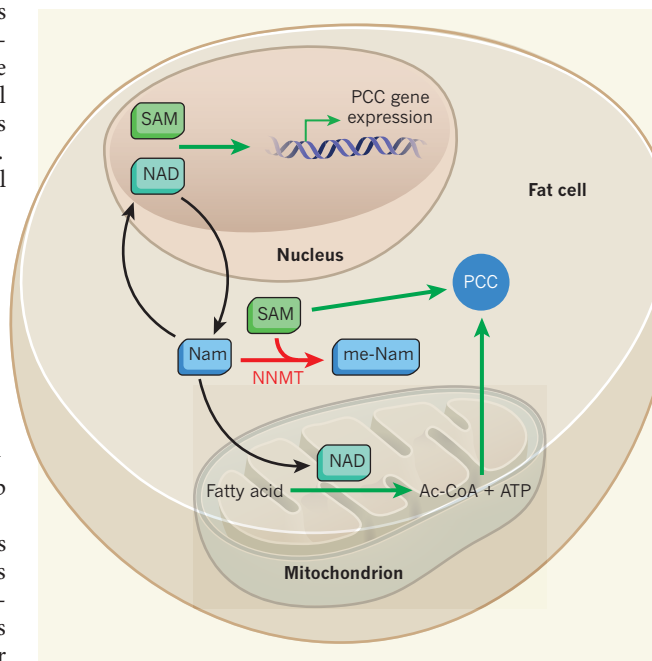
Figure 1 | Accelerating fat-cell metabolism. Kraus et al.[1] report that the enzyme nicotinamide *N*-methyltransferase (NNMT) is upregulated when fat levels increase, and that its inhibition prevents diet-induced obesity in mice. These effects are mediated by two cofactor molecules: nicotinamide adenine dinucleotide (NAD) and *S*-adenosine methionine (SAM). NAD consumption by enzymes produces nicotinamide (Nam), which can be resynthesized to NAD (salvage). However, NNMT adds a methyl group from SAM to nicotinamide (forming me-Nam), which prevents NAD salvage. NNMT inhibition therefore elevates levels of NAD and SAM, resulting in increased nuclear gene expression and cytoplasmic activity of the SAM-dependent polyamine catabolic cycle (PCC). PCC activation requires NAD-dependent fatty-acid oxidation and subsequent production of the molecules acetyl-CoA (Ac-CoA) and ATP in the mitochondrial cellular compartment. Pathways turned on by fat accumulation are shown in red. Those activated by NNMT inhibition are shown in green.

Nnmt was the most inversely correlated gene. Indeed, overexpression of *Nnmt* has been shown[5] to correlate with body mass index in a human group susceptible to diabetes.

Although it would be unsurprising if *Nnmt* expression were a consequence of chronic over-nutrition, rather than a cause of weight gain, Kraus et al. found that *Nnmt* expression in adipose tissue is correlated not only with the percentage of fat in 20 mouse strains, but also with sensitivity to diet-induced obesity in 25 strains. This qualified *Nnmt* as a candidate mediator of weight gain.

Kraus and colleagues developed a drug to inhibit the expression of *Nnmt* in a mouse strain prone to weight gain on a high-fat diet. This drug reduced *Nnmt* expression in fat and liver, but not elsewhere. Remarkably, the authors found that treated mice became resistant to diet-induced obesity. It is important to note, however, that mice have a much more rapid metabolism than humans, and that the trial was preventive in nature: the treatment enabled lean mice to resist weight gain and maintain insulin sensitivity on a high-fat diet, but obese mice were not examined for loss of fat with treatment.

What does NNMT do, and how does its inhibition reduce the incorporation of dietary fat into adipose tissue? Kraus and co-workers observed that drug-treated mice on a high-fat diet did not eat less, move more, produce more heat or excrete more fat than control animals. Instead, the animals' oxygen consumption was increased, suggestive of increased levels of fat oxidation (the process by which fat is converted to energy), and they exhibited increased activity of pathways that impair the ability of fat cells to store fat.

NNMT is involved in regulation of the cofactor nicotinamide adenine dinucleotide (NAD), a molecule needed in reactions that convert macronutrients into energy and that also acts as a consumed substrate of enzymes that perform many regulatory roles[6]. NAD consumption produces nicotinamide, a precursor of NAD; this creates a demand for nicotinamide to be resynthesized to NAD — a process termed salvage. NNMT attaches a methyl group to nicotinamide, preventing salvage[7]. The methyl group comes from a second cofactor, *S*-adenosine methionine (SAM), which also participates in diverse gene-regulatory and metabolic processes[8]. The researchers found that expression of *Nnmt* tended to lower NAD and SAM levels in fat, suggesting

that either or both cofactor-dependent hubs are functionally modulated by NNMT in fat cells (Fig. 1).

Although methylated nicotinamide is abundant in obesity and correlates with *Nnmt* expression, it does not seem to promote weight gain per se. In fact, Kraus and colleagues found that, at high doses, methylated nicotinamide inhibits NNMT, with activities similar to those of their *Nnmt*-targeting drug. The researchers showed that changes in metabolism in the drug-treated mice arose from NNMT inhibition through enhanced NAD- and SAM-dependent gene expression and enzyme activities. This is because, by inhibiting NNMT, there is more nicotinamide salvage to NAD, and SAM levels are preserved. Both cofactors regulate the state of histones, proteins in the nucleus that are responsible for chromosome packaging and tuning gene expression, and this may account for the gene-expression changes that the authors observed in drug-treated mice.

In the NNMT-inhibited mouse, more fat is oxidized and less is stored. NAD has two potential roles in this process. First, fatty-acid oxidation (which occurs in an energy-generating organelle, the mitochondrion) requires NAD. In adipose tissue, poor NAD salvage might directly retard this process, making more fat available for storage. Second, fatty-acid oxidation proceeds through an intermediate molecule, acetyl-CoA, and conditions of nutritional overload lead to build-up of this intermediate. This results in inhibition of mitochondrial enzymes, owing to transfer of acetyl groups from acetyl-CoA to the enzymes[2]. Although these modifications make mitochondria sluggish, it has been argued[2] that the ability to maximize weight gain in episodes of over-nutrition has been selected for throughout animal evolution. Because removal of mitochondrial acetyl groups depends on NAD, NNMT inhibition might reactivate mitochondrial-enzyme activities by increasing deacetylation, thereby improving oxygen consumption.

Although the NAD-dependent effects of NNMT inhibition in mitochondria remain largely unknown, Kraus *et al.* showed that, in the cytoplasm, the polyamine catabolic cycle, which is SAM-dependent, is accelerated by NNMT inhibition. This cycle drives oxygen consumption, because it uses up acetyl-CoA and the nucleotide ATP, thereby demanding that mitchondria oxidize fuels to produce more of these molecules[9,10]. A candidate activator of the cycle was unlikely to be identified by conventional means. Inhibition of NNMT seems to have done the trick.

The contributions of NAD- and SAM-dependent processes to adipocyte metabolism require further quantitative investigation. But there is little doubt that NNMT, an enzyme that is not only upregulated in obesity, but is also required for weight gain, will be tested as a target for weight loss. ∎

Charles Brenner *is in the Department of Biochemistry, the Obesity Research Initiative, University of Iowa, Iowa City, Iowa 52242, USA.*
e-mail: charles-brenner@uiowa.edu

1. Kraus, D. *et al. Nature* **508,** 258–262 (2014).
2. Ghanta, S., Grossmann, R. E. & Brenner, C. *Crit. Rev. Biochem. Mol. Biol.* **48,** 561–574 (2013).
3. Abel, E. D. *et al. Nature* **409,** 729–733 (2001).
4. Yang, Q. *et al. Nature* **436,** 356–362 (2005).
5. Lee, Y. H. *et al. Diabetologia* **48,** 1776–1783 (2005).
6. Bogan, K. L. & Brenner, C. *Annu. Rev. Nutr.* **28,** 115–130 (2008).
7. Aksoy, S., Szumlanski, C. L. & Weinshilboum, R. M. *J. Biol. Chem.* **269,** 14835–14840 (1994).
8. Lu, S. C. *Int. J. Biochem. Cell Biol.* **32,** 391–395 (2000).
9. Jell, J. *et al. J. Biol. Chem.* **282,** 8404–8413 (2007).
10. Pirinen, E. *et al. Mol. Cell. Biol.* **27,** 4953–4967 (2007).

QUANTUM PHYSICS

A strong hybrid couple

A single atom in an optical cavity is shown to interact strongly with an incoming photon and to switch the photon's state. This finding opens up a path towards optical quantum computation and quantum networks. SEE LETTERS P.237 & P.241

LUMING DUAN

When two torch beams cross, they go through each other without affecting one another. This means that the fundamental particles of light, photons, do not typically interact with each other, or with matter, unless designed to do so. Photons and matter interact when strong light beams propagate through a dense atomic medium, but this interaction is negligible for weak light pulses in a system of only a few atoms. In a future Internet based on the principles of quantum mechanics[1], which promises enhanced security and computational power, information will be carried by single photons, and the quantum state of those photons will need to be manipulated through their interaction with atoms. In this issue, Reiserer *et al.*[2] (page 237) and Tiecke *et al.*[3] (page 241) report independent experiments that bring this goal

a step closer. The researchers have designed systems in which a single atom switches the state of a single photon contained in a faint light pulse.

The experiments represent the culmination of decades of research into atom–photon coupling in an optical cavity[1]. The Fabry–Perót version of an optical cavity consists of two highly reflective mirrors between which a photon bounces many times. This arrangement allows an atom trapped inside the cavity to be strongly coupled with the photon. Reiserer *et al.* used a Fabry–Perót cavity in which one of the mirrors has a significantly higher reflectivity, and thus lower transmissivity, than the other, so that a photon enters and leaves the cavity mainly through the lower-reflectivity mirror. Tiecke *et al.* designed a special type of cavity known as a photonic crystal cavity, which has the same function as Reiserer and colleagues' Fabry–Perót cavity but has a tiny
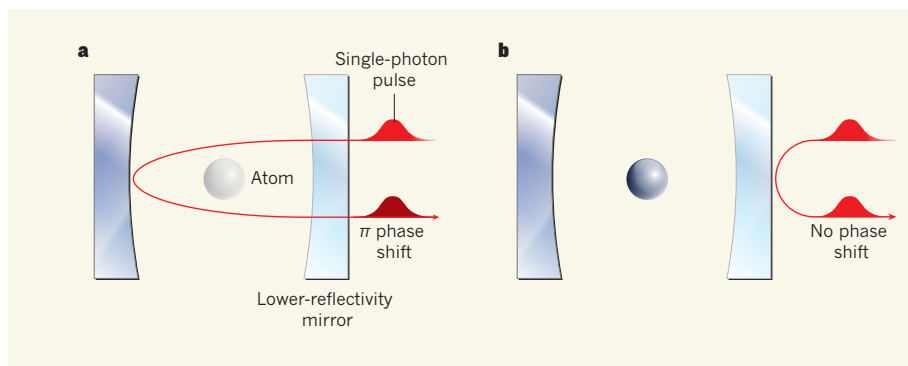
Figure 1 | An optical switch. Reiserer *et al.*[2] and Tiecke *et al.*[3] have designed systems in which a single atom trapped in an optical cavity, here formed by an arrangement of two mirrors of different reflectivity, switches the state of a photon in an incoming light pulse. **a,** If the atom is in a quantum state that does not couple with the cavity, equivalent to there being no atom in the device, a single-photon pulse resonant with one of the cavity's optical modes of oscillation will enter the cavity through the lower-reflectivity mirror and leave it with a phase shift of π radians, illustrated by a darker red than that of the original pulse. **b,** If the atom is in a state that couples with the cavity, it will shift the frequency of the cavity's mode and the pulse will now be off-resonant with it. Therefore, the pulse will not enter the cavity and will bounce back with no phase shift.

cavity volume, which helps to further enhance the atom–photon coupling.

To switch the photon's state using a single atom, the two studies exploited a scheme proposed for optical quantum computation[4]. To understand this scheme, consider a spring fixed at either end. Just as the spring carries vibrations of only certain frequencies, an optical cavity in which the mirrors are separated by a fixed distance allows light of only certain frequencies, called cavity modes, to enter it. When an incoming single-photon pulse has the same frequency as one of the cavity modes (in other words, the pulse is resonant with the cavity), it enters the cavity and leaves it through the same mirror (Fig. 1a). In this process, the photon pulse, as a wave, undergoes a shift in phase of π radians; the phase quantifies a wave's local amplitude as it oscillates between its minimum and maximum values. Now, if there is an atom inside the cavity, it couples with the cavity modes and shifts their frequencies. Because of this shift, the incoming pulse that was resonant with the cavity in the absence of the atom will no longer be resonant with it, and thus will not enter the cavity. As a result, it bounces directly back, with no phase shift (Fig. 1b).

In the authors' experiments, such a conditional phase shift is achieved by preparing the atom in two quantum states: one that couples with the cavity ('presence state') and another that does not ('absence state'). The researchers then prepared the atom in a quantum-mechanical superposition of presence and absence states. Together with the conditional phase shift, this superposition state allowed them to implement a quantum logic gate — the basic building block of quantum computation — between the atom and the photon. Such a gate is crucial for creating quantum networks in which information is stored in, and retrieved from, atoms and transmitted to distant locations by means of single-photon pulses.

Reiserer and colleagues went on to demonstrate that the gate generates entanglement between the atom and the photon, and that this quantum effect mediates an interaction between different single-photon pulses. By shining a succession of resonant single-photon pulses on the cavity, the authors showed that successively reflected pulses become quantum entangled. This mediated interaction offers a means to realize quantum logic gates between single-photon pulses, providing a scalable platform for optical quantum computation[4].

The demonstrated conditional π phase shift also has applications in nonlinear quantum optics. Tiecke et al. observed a nonlinear response from the cavity in which the single-photon and the two-photon components of an incoming light pulse are routed to different paths. With reduced photon loss and improved localization of the atom inside the cavity, the conditional phase shift can also be used to prepare a Schrödinger's cat state — an

intriguing quantum superposition of classically distinct states — for the reflected light pulse[5]. Finally, the physical principles behind these experiments are not limited to atom–photon coupling. They could be applied to coupling between a quantum dot (an artificial atom) and infrared photons, or between a superconducting quantum bit and microwave photons. ∎

Luming Duan is in the Physics Department, University of Michigan, Ann Arbor, Michigan 418003, USA, and in the Center for Quantum Information, Tsinghua University, Beijing, China.
e-mail: lmduan@umich.edu

1. Kimble, H. J. *Nature* **453,** 1023–1030 (2008).
2. Reiserer, A., Kalb, N., Rempe, G. & Ritter, S. *Nature* **508,** 237–240 (2014).
3. Tiecke, T. G. *et al. Nature* **508,** 241–244 (2014).
4. Duan, L.-M. & Kimble, H. J. *Phys. Rev. Lett.* **92,** 127902 (2004).
5. Wang, B. & Duan, L.-M. *Phys. Rev. A* **72,** 022320 (2005).

The purple heart of photosynthesis

The structure of a photosynthetic complex from a purple bacterium reveals a new class of light-harvesting protein and the channels that might allow electron-transporting molecules to escape this otherwise closed system. SEE ARTICLE P.228

RICHARD J. COGDELL & ALEKSANDER W. ROSZAK

Photosynthesis is one of the most fundamentally important biological processes on Earth. It provides all the oxygen we breathe, the food we eat and the fossil fuels we so greedily consume. Purple photosynthetic bacteria have proved to be excellent model organisms with which to study the reactions that occur at the early stages of photosynthesis. In this issue, Niwa et al.[1] (page 228) report the X-ray crystal structure of part of the photosynthetic machinery from the bacterium *Thermochromatium tepidum*, at near-atomic

resolution. Their findings greatly advance our understanding of the detailed molecular mechanisms involved in primary photosynthetic reactions.

The pigment–protein complexes that catalyse these reactions are located in and on intracytoplasmic membranes — infoldings of the cell membrane. The major light-absorbing pigments in purple bacteria are bacteriochlorophyll and carotenoid molecules (the orange pigments found in carrots). These pigments are non-covalently bound to well-defined membrane proteins known as the light-harvesting complexes (LH1 and LH2) and reaction centres (RCs). In most species of purple bacterium,
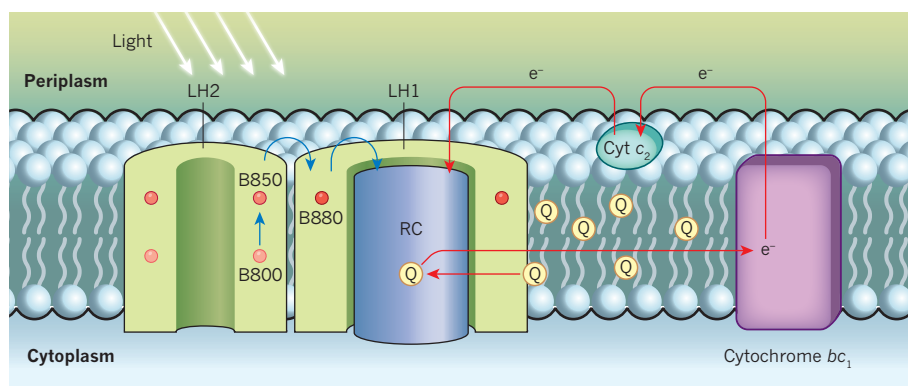
Figure 1 | Typical organization of a photosynthetic unit in the intracellular membrane of purple bacteria. Light-harvesting complex 2 (LH2) contains two groups of bacteriochlorophyll (BChL) pigments, B800 and B850. Light energy absorbed by B800 is transferred to B850 (blue arrows). It is then passed to another group of BChLs (B880) in light-harvesting complex 1 (LH1), and finally into the reaction centre (RC). The excited RC fully reduces ubiquinone molecules (Q), which leave the LH1–RC complex and pass into the membrane, transferring electrons (e-) to the cytochrome bc_1 complex; this transfer forms part of a cyclic electron-transport pathway (red arrows) that drives photosynthesis. This pathway is completed by a soluble protein, cytochrome c_2 (Cyt c_2). The periplasm is the region between the cell membrane and the cell wall of the bacterium.
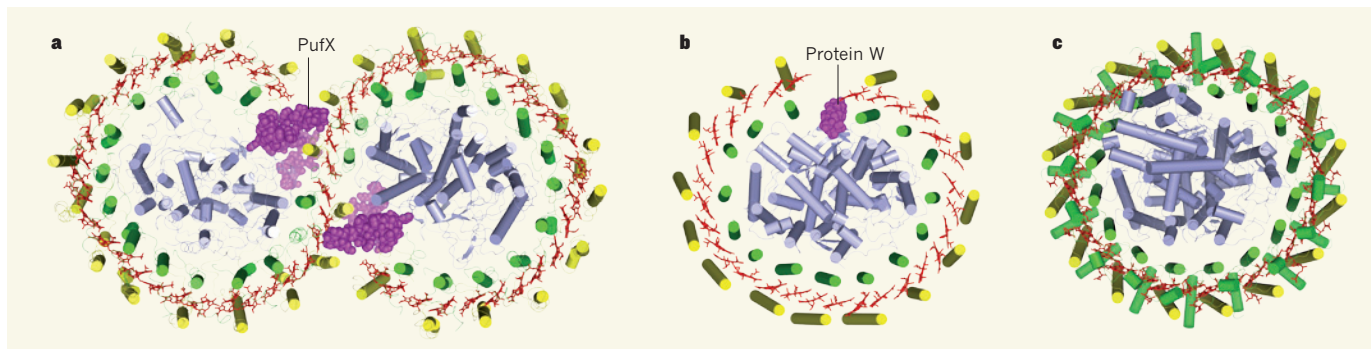
Figure 2 | X-ray crystal structures of three types of the LH1–RC complex. **a**, A model of the dimeric RC–LH1–PufX complex from *Rhodobacter sphaeroides*[5] (8 ångströms resolution), in which PufX proteins create gaps in the LH1 rings. Yellow and green cylinders represent helical portions of protein subunits of LH1 (α- and β-apoproteins, respectively); blue cylinders are helices of the RC. Non-helical portions of protein chains are shown as thin coils. Bacteriochlorophyll pigments of the LH1 complex are shown in red. No LH1 carotenoid pigments or RC cofactors are shown, for clarity. **b**, A model of the monomeric RC–LH1–protein W complex from *Rhodopseudomonas palustris*[6] (4.8 Å resolution). Here, protein W creates a gap in the LH1 ring. **c**, Niwa *et al.*[1] report this 3-Å model of the monomeric RC–LH1 complex from *Thermochromatium tepidum*. In the absence of a protein such as PufX or protein W, there is no visible break in the LH1 ring.

light energy is initially absorbed by LH2 and then transferred by LH1 to the RC, in which it is used to drive the primary redox reactions of photosynthesis[2] (Fig. 1). These reactions involve the conversion of ubiquinone molecules to their completely reduced form, which carries electrons to the cytochrome bc_1 complex as part of a cyclic electron-transport pathway.

The value of this model system is clearly illustrated by two landmark studies. The first determined the X-ray crystal structure of the RC from the bacterium *Blastochloris viridis*[3]; this was the first crystal structure of a membrane protein to be determined, and led to the award of the Nobel Prize in Chemistry in 1988. The second study solved the crystal structure of the LH2 complex from the bacterium *Rhodopseudomonas acidophila*[4]. These two structures revealed fundamental principles that underlie our current understanding of the energy-transfer processes of photosynthesis. Niwa and colleagues' study describes the structure of the *T. tepidum* LH1–RC complex, and promises to be just as influential as the two earlier papers.

Earlier low-resolution structures had revealed that LH1–RC complexes can occur as dimers in *Rhodobacter sphaeroides*[5] and as monomers in *Rhodopseudomonas palustris*[6] (Fig. 2a, b). In these bacteria, the basic structures are elliptical and contain gaps in the LH1 ring. It has been suggested that reduced ubiquinone passes through these gaps to leave the RC and connect to the cytochrome bc_1 complex. In both of these structures, the gaps are introduced by a protein: PufX in the dimeric complex and protein W in the monomeric complex.

Niwa and co-workers' structure shows that a third class of LH1–RC complex is possible. In their case, the LH1 ring is more uniformly elliptical than the other two structures, but there is no analogue of PufX or protein W (Fig. 2c). Nevertheless, channels are present at a level corresponding to the middle of the transmembrane region of their complex (see Fig. 5 of the paper[1]). The authors suggest that

these channels provide the same function as the gaps in the previously reported structures, allowing reduced ubiquinone molecules to connect with the cyclic electron-transport pathway. Their structure confirms conclusions from previous molecular-dynamic simulations that also suggested this type of mechanism for ubiquinone diffusion through the LH1 ring[7,8]. This escape of the reduced ubiquinone from the 'core' complex is essential for cyclic electron transport in purple bacteria; without it, they cannot grow through photosynthesis[9].

Thermochromatium tepidum was discovered in hot springs, and has an optimum growth temperature of 50 °C. Its photosynthetic complex is therefore more stable at higher temperatures than the complexes of other purple bacteria. Niwa *et al.* describe in detail the amino-acid residues and bonds involved in binding calcium ions to the carboxy termini of the LH1 protein subunits, and suggest that these interactions are crucial for the increased thermostability of the complex.

Another interesting quirk of the LH1–RC complex in *T. tepidum* is that the longest-wavelength band (called the Q_y band) in the near-infrared absorption spectrum of bacteriochlorophyll is considerably redshifted compared with that of the other complexes shown in Figure 2. The energy transfer from LH1 to RC takes place from this Q_y band. The authors conclude that the calcium-ion binding observed in their structure, and the effects that this has on the bacteriochlorophyll molecules in LH1, causes this large redshift.

Niwa and colleagues' structure also shows, for the first time in an LH1 complex, the exact position and orientation of the carotenoid molecules; in this case the carotenoid is spirilloxanthin. Carotenoids are required for both light harvesting and photoprotection — in their absence, illumination in the presence of oxygen destroys light-harvesting complexes and kills the photosynthetic organism in which they reside. Niwa *et al.* observe

that the end of each spirilloxanthin molecule in the LH1 complex is in close proximity to a histidine amino-acid residue that is bound to the central magnesium atom of a bacteriochlorophyll within the same protein subunit. In LH1 complexes from other species of purple bacterium, removal of this carotenoid results in a blueshift of the bacteriochlorophyll Q_y absorption band[10]. Loss of the carotenoid interactions described above probably explains this long-standing observation.

The report of a new structure of a membrane protein is always an event. That is particularly true in this case, when the protein involved is at the heart of a process as vital as photosynthesis. It will be exciting to see if comparing theoretical modelling of the spectroscopic properties and energy-transfer reactions of this LH1–RC complex, based on Niwa and co-workers structure, with experimental measurements produces a consistent, detailed mechanistic understanding of photosynthesis. ∎

Richard J. Cogdell *is at the Institute of Molecular, Cell and Systems Biology, University of Glasgow, Glasgow G12 8QQ, UK.* **Aleksander W. Roszak** *is in the WestCHEM, School of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK.*
e-mails: richard.cogdell@glasgow.ac.uk; aleksander.roszak@glasgow.ac.uk

1. Niwa, S. *et al.* *Nature* **508,** 228–232 (2014).
2. Cogdell, R. J., Gall, A. & Köhler, J. *Q. Rev. Biophys.* **39,** 227–324 (2006).
3. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. *Nature* **318,** 618–624 (1985).
4. McDermott, G. *et al.* *Nature* **374,** 517–521 (1995).
5. Qian, P. *et al.* *Biochemistry* **52,** 7575–7585 (2013).
6. Roszak, A. W. *et al.* *Science* **302,** 1969–1972 (2003).
7. Aird, A., Wrachtrup, J., Schulten, K. & Tietz, C. *Biophys. J.* **92,** 23–33 (2007).
8. Walz, T. & Ghosh, R. *J. Mol. Biol.* **265,** 107–111 (1997).
9. Barz, W. P. *et al.* *Biochemistry* **34,** 15235–15247 (1995).
10. Sistrom, W. R., Griffiths, M. & Stanier, R. Y. *J. Cell. Comp. Physiol.* **48,** 473–515 (1956).

ARTICLE

Transcriptional landscape of the prenatal human brain

Jeremy A. Miller[1]*, Song-Lin Ding[1]*, Susan M. Sunkin[1], Kimberly A. Smith[1], Lydia Ng[1], Aaron Szafer[1], Amanda Ebbert[1], Zackery L. Riley[1], Joshua J. Royall[1], Kaylynn Aiona[1], James M. Arnold[1], Crissa Bennet[1], Darren Bertagnolli[1], Krissy Brouner[1], Stephanie Butler[1], Shiella Caldejon[1], Anita Carey[1], Christine Cuhaciyan[1], Rachel A. Dalley[1], Nick Dee[1], Tim A. Dolbeare[1], Benjamin A. C. Facer[1], David Feng[1], Tim P. Fliss[1], Garrett Gee[1], Jeff Goldy[1], Lindsey Gourley[1], Benjamin W. Gregor[1], Guangyu Gu[1], Robert E. Howard[1], Jayson M. Jochim[1], Chihchau L. Kuan[1], Christopher Lau[1], Chang-Kyu Lee[1], Felix Lee[1], Tracy A. Lemon[1], Phil Lesnar[1], Bergen McMurray[1], Naveed Mastan[1], Nerick Mosqueda[1], Theresa Naluai-Cecchini[2], Nhan-Kiet Ngo[1], Julie Nyhus[1], Aaron Oldre[1], Eric Olson[1], Jody Parente[1], Patrick D. Parker[1], Sheana E. Parry[1], Allison Stevens[3,4], Mihovil Pletikos[5], Melissa Reding[1], Kate Roll[1], David Sandman[1], Melaine Sarreal[1], Sheila Shapouri[1], Nadiya V. Shapovalova[1], Elaine H. Shen[1], Nathan Sjoquist[1], Clifford R. Slaughterbeck[1], Michael Smith[1], Andy J. Sodt[1], Derric Williams[1], Lilla Zöllei[3], Bruce Fischl[3,4], Mark B. Gerstein[6,7], Daniel H. Geschwind[8], Ian A. Glass[2], Michael J. Hawrylycz[1], Robert F. Hevner[9,10], Hao Huang[11], Allan R. Jones[1], James A. Knowles[12], Pat Levitt[13,14], John W. Phillips[1], Nenad Šestan[5], Paul Wohnoutka[1], Chinh Dang[1], Amy Bernard[1], John G. Hohmann[1] & Ed S. Lein[1]

The anatomical and functional architecture of the human brain is mainly determined by prenatal transcriptional processes. We describe an anatomically comprehensive atlas of the mid-gestational human brain, including _de novo_ reference atlases, _in situ_ hybridization, ultra-high-resolution magnetic resonance imaging (MRI) and microarray analysis on highly discrete laser-microdissected brain regions. In developing cerebral cortex, transcriptional differences are found between different proliferative and post-mitotic layers, wherein laminar signatures reflect cellular composition and developmental processes. Cytoarchitectural differences between human and mouse have molecular correlates, including species differences in gene expression in subplate, although surprisingly we find minimal differences between the inner and outer subventricular zones even though the outer zone is expanded in humans. Both germinal and post-mitotic cortical layers exhibit fronto-temporal gradients, with particular enrichment in the frontal lobe. Finally, many neuro-developmental disorder and human-evolution-related genes show patterned expression, potentially underlying unique features of human cortical formation. These data provide a rich, freely-accessible resource for understanding human brain development.

The human brain develops following a complex, highly stereotyped series of histogenic events that depend on regulated differential gene expression, and acquired or inherited disruption can lead to devastating consequences. Mainly due to limitations in access to human pre-natal tissue, most developmental studies are performed in mouse or non-human primates (however, see refs 1–5). Yet significant species differences exist, necessitating the study of human brain. For example, the human neocortex has undergone massive evolutionary expansion, particularly in superficial layers, probably due to differences in rates of progenitor pool expansion during neurogenesis compared to other species[6]. A secondary progenitor zone, the subventricular zone (SZ) is present in all mammals, but is split into an outer and inner region in primates[7]. The transient subplate zone (SP) is greatly expanded in human[8], as is the subpial granular zone (SG), a transient compartment at the pial surface composed primarily of tangentially migrating neurons[9]. Furthermore, there is evidence for species differences in the developmental origin of cortical GABAergic interneurons. In mouse, nearly all originate from the striatal ganglionic eminences (GEs) of the ventral telencephalon[10]; however, the origin of human cortical interneurons remains controversial[7,11–14]. Finally, understanding the emergence of cortical specialization for language can only be studied in humans.

Recent studies have begun to analyse the developing brain and neo-cortical transcriptome. Profiling of mid-gestational human brain[3,5] identified many genes differentially expressed between major regions, including genes associated with human-accelerated conserved noncoding sequences (haCNSs)[15]. Gene expression also varies between cell populations, and more detailed analysis of layers of fetal mouse neocortex found >2,500 genes differentially expressed between ventricular zone (VZ), subventricular zone, intermediate zone (IZ), and cortical plate (CP)[16]. Species differences in distinct fetal transient zones, including the subventricular zone[2], subplate zone[17,18], cortical plate[6], and subpial granular zone[9], have also been described.

The goal of the current project was to create resources for studying prenatal human brain development and the early roots of neurodevelopmental

[1]Allen Institute for Brain Science, Seattle, Washington 98103, USA. [2]Division of Genetic Medicine, Department of Pediatrics, University of Washington, 1959 North East Pacific Street, Box 356320, Seattle, Washington 98195, USA. [3]Department of Radiology, Harvard Medical School, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. [4]Computer Science and AI Lab, MIT, Cambridge, Massachusetts 02139, USA. [5]Department of Neurobiology and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, Connecticut 06510, USA. [6]Program in Computational Biology and Bioinformatics, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA. [7]Department of Computer Science, Yale University, New Haven, Connecticut 06520, USA. [8]Program in Neurogenetics, Department of Neurology and Semel Institute David Geffen School of Medicine, UCLA, Los Angeles, California 90095, USA. [9]Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, Washington 98101, USA. [10]Department of Neurological Surgery, University of Washington School of Medicine, Seattle, Washington 98105, USA. [11]Advanced Imaging Research Center, UT Southwestern Medical Center, Dallas, Texas 75390, USA. [12]Zilkha Neurogenetic Institute, and Department of Psychiatry, University of Southern California, Los Angeles, California 90033, USA. [13]Department of Pediatrics, Children's Hospital, Los Angeles, California 90027, USA. [14]Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA.
*These authors contributed equally to this work.

and psychiatric disorders[19]. These include anatomical reference atlases similar to those for model organisms[20–22], and an anatomically comprehensive, detailed transcriptional profiling of normal mid-gestational brain modelled on atlases of adult mouse and human brain[23,24] and using methods for selective analysis of discrete structural nuclei and layers[25]. These data are freely accessible as part of the BrainSpan Atlas of the Developing Human Brain (http://brainspan.org) via the Allen Brain Atlas data portal (http://www.brain-map.org).

Transcriptome analysis of prenatal brain

Four intact high-quality mid-gestational brains, two from 15 and 16 post-conceptual weeks (pcw) and two from 21 pcw (Supplementary Table 1), were used to create detailed *de novo* reference atlases and transcriptome data sets (Fig. 1). The entire left hemisphere of each specimen was coronally, serially cryosectioned onto polyethylene naphthalate (PEN) membrane slides for laser microdissection (LMD), with interleaved slides for histological staining (Nissl, acetylcholinesterase (AChE), and *in situ* hybridization for *GAP43*) for detailed structure identification. Approximately 300 anatomical regions per specimen were isolated for RNA isolation, amplification and microarray analysis on custom 64K Agilent microarrays[23] (Fig. 1d, Supplementary Table 2 and Extended Data Fig. 1). For one 15 pcw and one 21 pcw specimen, the right hemisphere was processed similarly but used for *in situ* hybridization (Fig. 1b) and Nissl staining. These data were anatomically delineated to make digital reference atlases (Fig. 1a), which allow the representation of transcriptome data in native anatomical coordinates. Figure 1e illustrates the specificity of anatomical sampling using this representation. For example, the folate receptor *FOLR1* is selectively expressed in the ventricular zone and ganglionic eminences. Sufficient folate intake is essential for normal neuronal development[26], and mutations in *FOLR1* cause severe neurological sequelae due to cerebral folate transport deficiency. Similarly, two genes associated with abnormal cortical development in holoprosencephaly, *TGIF1* and *SIX3* (ref. 27), are also enriched in cortical germinal zones. Finally, structural magnetic resonance imaging (MRI) and diffusion weighted MRI (DWI) data from 3 approximately age-matched brains (Fig. 1c and Extended Data Fig. 2), as well as reconstruction of fetal white matter tracts for three additional brains (Extended Data Fig. 3 and see ref. 28), are included for anatomical reference in the online resource.

Laminar transcriptional patterning

We assayed ~25 areas of the developing neocortex, delineating nine layers per area (here referring to fetal mitotic and post-mitotic zones and not layers 1–6 of mature neocortex)[29]: Subpial granular zone, marginal zone (MZ), outer and inner cortical plate (CPo; CPi), subplate zone, intermediate zone (or inner subplate), outer and inner subventricular zone (SZo; SZi), and ventricular zone (Fig. 2a). Approximately 95% of RefSeq genes are expressed in developing neocortex, compared with 84% identified using these arrays in adult[23]. Different layers show robust and unique molecular signatures, and samples group by layer using hierarchical clustering (using differential genes from ANOVA, $P < 10^{-29}$) at both time points (Fig. 2b). Samples also cluster with multidimensional scaling, where dimension one separates samples by layer, with germinal zones (VZ, SZi, SZo) distinct from layers containing primarily post-mitotic cells, and dimension two roughly reflects rostro-caudal position (Fig. 2c).

To identify laminar signatures at 21 pcw, we correlated each gene with a binary vector (1 in tested layer versus 0 elsewhere), which identified ~2,000 layer-enriched genes with $R > 0.5$ in both brains (Supplementary Table 3). Each layer included genes with high laminar specificity (Fig. 2d), although the SZi profiles tended to overlap with the neighbouring SZo and ventricular zone. *In situ* hybridization validated the specificity of layer enriched genes (Fig. 2e). For example, the Cajal-Retzius cell marker *CALB2* (ref. 30) showed enrichment in the marginal zone and subpial granular zone as expected. The cortical progenitor markers *TBR2* and *PAX6* are enriched in germinal layers as in mouse, although
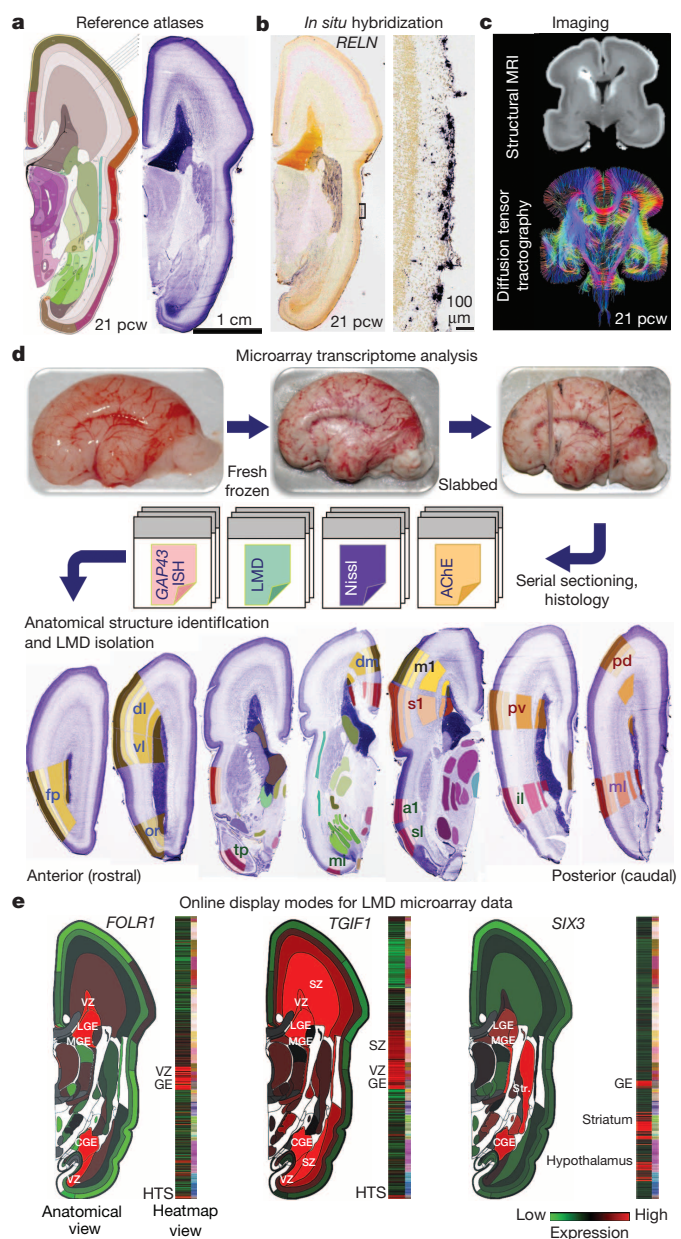
Figure 1 | Prenatal human brain atlas components. a, Nissl stained (right) and corresponding annotated reference atlas (left) plate, colour coded by structure. **b**, *In situ* hybridization for *RELN* showing expression in Cajal-Retzius cells at low (left) and high (right) magnification in marginal zone. **c**, High resolution MRI and tract DWI of fixed *ex cranio* brain. **d**, Experimental strategy for systematic histology, anatomical delineation and LMD-based isolation of discrete anatomical structures for microarray analysis. Nissl, acetylcholinesterase (AChE) and *GAP43 in situ* hybridization were used to identify structures. **e**, Quantitative representation of microarray data for *FOLR1*, *TGIF1* and *SIX3*. GE, ganglionic eminence. See Supplementary Table 2 for other anatomical abbreviations.

Pax6 in mouse is restricted to the ventricular zone, whereas it is also highly expressed in the human subventricular zone[31]. Ventricular-zone-restricted *GFAP* expression probably marks radial glia (RG)[32]. Finally, expression of ZIC1, associated with Dandy-Walker congenital brain malformation[33], was restricted to the pia mater overlying the cortex, therefore indicating that subpial granular zone samples captured pial cells in addition to granule cells. However, whereas mouse *Zic1* is expressed by virtually all Cajal-Retzius neurons[34], our results indicate that this is not true in human as ZIC1 and CALB2 expression do not overlap in the marginal zone (Figs 1b and 2e).

Figure 2 | Laminar gene expression mirrors developmental processes in prenatal neocortex. a, Nissl section from 16 pcw cortex showing layers dissected for analysis (represented by the colour bar). The outer and inner fibre layers (OFL, IFL) were omitted from the dissection. **b**, Sample clustering based on the 100 most significant genes differentiating layers by ANOVA ($P < 10^{-29}$) groups samples by layer at both time points. **c**, Multidimensional scaling using all genes demonstrates clustering of post-mitotic versus germinal zones (dimension 1) and to a lesser extent by rostrocaudal position (dimension 2). **d**, Layer-enriched gene expression, based on correlation to binary templates at 21 pcw (Methods). Enriched genes in each layer relate to cellular makeup and developmental maturity of those cells. Asterisks indicate genes investigated further (see panel **e**). P values are included in parentheses. Additional categories just missing significance at $P < 0.05$ are shown in grey. **e**, Validation of laminar enrichment by *in situ* hybridization for genes with asterisks in **d**. See Supplementary Table 2 for anatomical abbreviations.
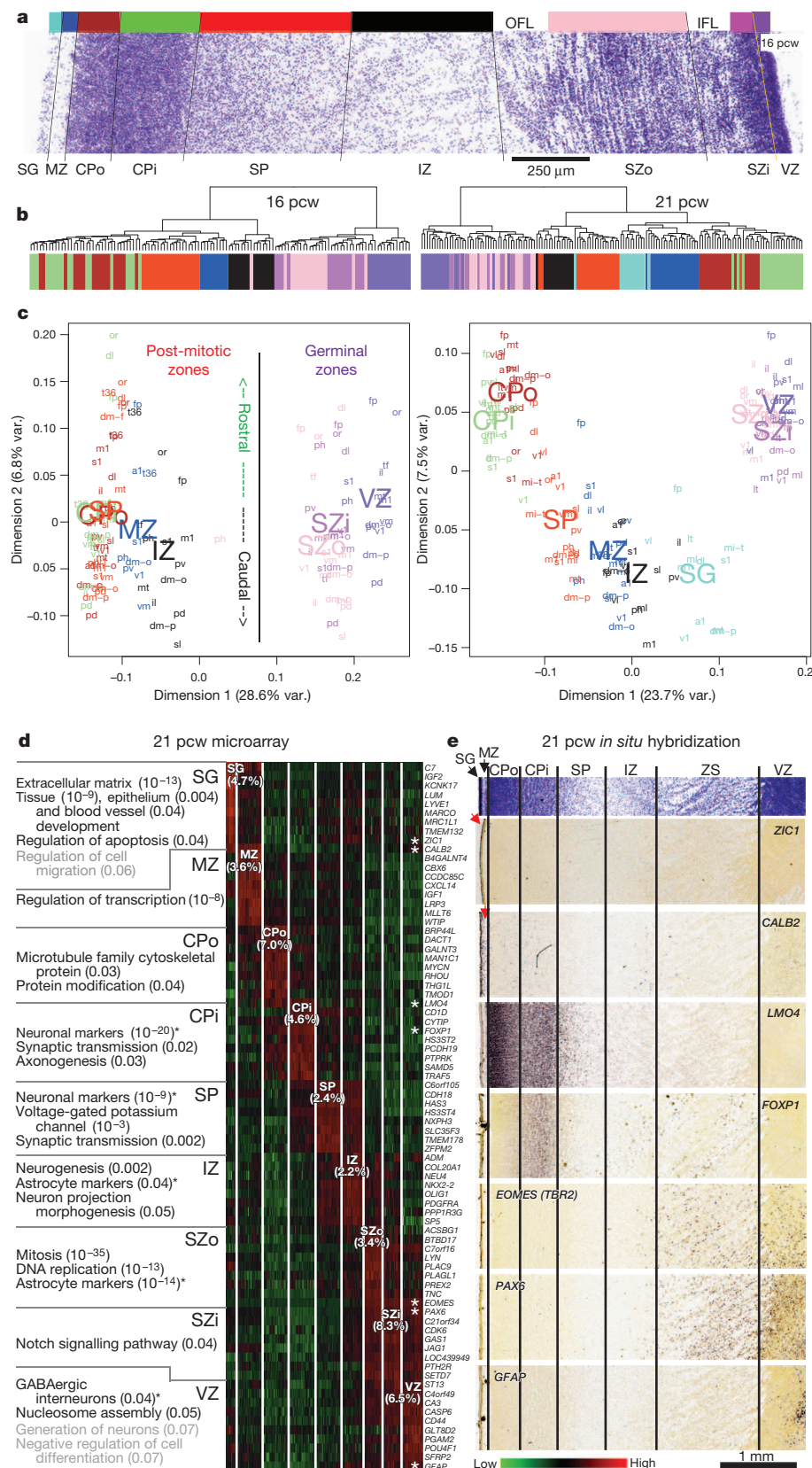
These laminar expression patterns mirror cellular composition and developmental processes, shown by enrichment analysis (Fig. 2d, Supplementary Table 4 and Methods). SZo-enriched categories primarily related to cell division and contained many astrocytic markers probably expressed in outer radial glia (ORG)[7]. Functional ontology of post-mitotic layers reflected developmental maturity. Subplate, which contains the earliest-generated neurons, showed enrichment for mature neuronal markers and synaptic transmission, reflecting early thalamic

afferent input by mid-gestation[35]. The next oldest neurons in CPi are additionally enriched for genes involved in forming connections, whereas the youngest neurons in CPo are primarily enriched for terms related to metabolism rather than mature neuronal function.

Gene networks discriminate fetal cell types

To identify principal features of the developing cortical transcriptome, we performed weighted gene co-expression network analysis (WGCNA)[36] on all 526 neocortical samples, and identified 42 modules of co-expressed genes (Fig. 3a, Supplementary Tables 5 and 6 and Supplementary Methods). WGCNA clusters genes with similar expression patterns in an unbiased manner, allowing a biological interpretation of transcriptional patterns (layer, cell type, biological process, disease and so on)[23,36–38]. Here, most gene clusters ('modules') corresponded to layers and/or changes with age, (Fig. 3a, b and Extended Data Fig. 4) whereas areal patterning appeared to be a more subtle transcriptional feature. For example, module C16 is enriched in subplate (Fig. 3b, lower right), and shows hallmarks of mature neuronal function. Module C38 contains genes enriched in germinal layers, and also decreased expression with age (Fig. 3b, upper left). This module has a large signature of glia and cell division, suggesting that these genes reflect decreasing progenitor cell division. Conversely, module C22 is enriched in newly generated post-mitotic neurons of the cortical plate, and increases with age (Fig. 3b, lower left). Importantly, given the small sample size, this temporal patterning in C38 and C22 is corroborated by RNA-seq data from a larger time series of cortical development contained in the BrainSpan resource (Extended Data Fig. 5). Genes in module C22 significantly overlap with genes showing altered expression in postnatal human brains of patients affected by autism[38]. This suggests involvement of autism risk factors in early development of excitatory cortical neurons, consistent with other recent studies[39,40].

Finally, we identified a module (C31) with particular enrichment in subpial granular zone and ventricular zone (Fig. 3b, upper right), containing many interneuron-associated genes (Fig. 3c). DLX1 and DLX2, homeobox transcription factors essential for interneuron migration and

survival[10], were central ('hub') genes therein. There is controversy regarding the origin of cortical interneurons in primates. The argument has been made that a substantial proportion of interneurons are generated locally in the ventricular zone[12,13]. However, several recent studies have shown strong evidence that, as with rodents, primate interneurons are generated extracortically in the ganglionic eminences[11,14]. To address this issue, we generated a new network using neocortical as well as ganglionic eminences samples, and examined the distribution of C31 genes therein. Most genes were assigned to a module showing common enrichment in both the ganglionic eminences and the ventricular zone (Extended Data Fig. 6). Although this finding does not resolve the origin of cortical interneurons, it shows that transcriptional programs associated with these cells in both structures are highly similar.

Germinal layers contain various cortical progenitors including radial glia in the ventricular zone, intermediate progenitors (IP) in the SZi, and ORG in the SZo (reviewed in ref. 41), and these radial glia may be quite diverse[42]. To search for coherent expression profiles marking putative progenitor populations, we created a consensus co-expression network using samples from VZ, SZi and SZo in the 15 and 16 pcw brains, only including genes differentially expressed between these layers (Fig. 3d; Supplementary Table 7 and Supplementary Methods). This network will only identify intergenic relationships common to both brains and should therefore produce robust co-expression relationships that exclude specimen-specific features or changes with age. We found eight co-expression modules with selective enrichment in either the subventricular zone or the ventricular zone (Fig. 3d, e), highly conserved between brains ($P < 10^{-50}$), and highly distinct. Both ventricular-zone-enriched modules (G7, G8) contained cell cycle genes and many astrocyte markers, indicating that these modules may represent different radial glia populations. Fluorescent in situ hybridization (FISH) confirmed that representative genes in both G7 (NR2E1) and G8 (SPATA13) are enriched in the ventricular zone versus subventricular zone (Fig. 3f and Extended Data Fig. 7). Surprisingly, rather than labelling distinct cell populations in the ventricular zone, these genes labelled mutually exclusive
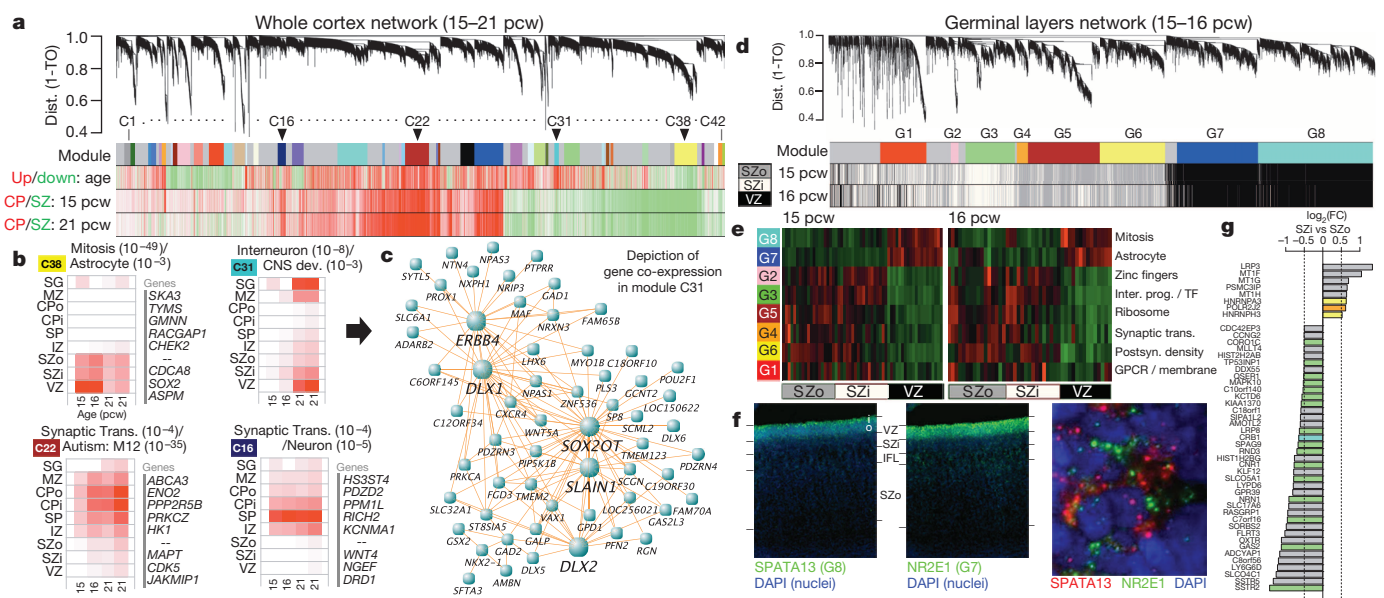
Figure 3 | Co-expression analyses of prenatal cortex. a, WGCNA cluster dendrogram on all 526 neocortical samples groups genes into 42 distinct modules (row 1). Rows 2–4 show strong differential expression relationships between module genes and age or cortical layer. TO, topological overlap. **b**, Module eigengene (ME) expression of four notable modules in **a**, averaged across brain and layer. Modules are biologically characterized with significant category enrichments and representative genes. White, low expression; red, high expression. **c**, Many top gene–gene connections for module C31 are shown, including several known GABAergic interneuron genes. **d**, Cluster dendrogram for consensus network focused on germinal layers identifies modules (row 1) enriched in each layer (rows 2 and 3). **e**, ME heatmap shows differential VZ/SZ expression for 8 modules, along with enriched gene sets. **f**, FISH on 15 pcw frontal cortex shows enrichment of SPATA13 and NR2E1 in mutually exclusive subcellular localization in VZ. **g**, Genes enriched in SZ and differentially expressed between SZo and SZi. Genes colour-coded by module (grey = unassigned).

subcellular locations within the same ventricular zone cells (note non-overlapping cytoplasmic localization of SPATA13 and NR2E1), indicating that these modules probably represent differentially regulated biological processes within radial glia cells.

We were particularly interested in identifying differences between the inner and outer subventricular zone (SZi and SZo), which is absent in mice. Modules G2 and G3 are enriched in the SZi (Fig. 3d, white bars in bottom two rows), and G3 is enriched for genes marking *Svet1+* intermediate progenitors (IPs) in E14 mouse subventricular zone[43], including *ELAVL4, LRP8, NEUROD1, NRN1, SLC17A6, SSTR2* and *TP53INP1*. Modules G4–G6 are enriched in the SZo (Fig. 3d, grey bars); however, these modules are enriched for neuron-associated categories probably expressed in post-mitotic cells during their radial migration. To identify expression specific to SZo or SZi using a more targeted approach, we searched for genes both maximally expressed in the subventricular zone and differentially expressed between SZo and SZi (Fig. 3g; $t > 4$, $P < 0.01$, $\log_2(FC) > 0.5$). Remarkably, few genes met both criteria at 15 or 16 pcw: 39 genes were enriched in SZi (including several in module G3), whereas only eight were specifically enriched in SZo.

These results are consistent with a previous study of laminar enrichment in 13–16 pcw human prenatal cortex, which also found few genes specifically enriched in SZo (55) or SZi (61)[2].

Species differences in subplate

The subplate is a largely transient zone beneath the cortical plate that plays an important role in establishment of thalamocortical connectivity (see ref. 35 for review). Subplate generation is protracted in primates[8], and its thickness particularly expanded in human[8]. In mouse and other species[18] this layer is molecularly distinct, and our laminar profiling also identified many subplate-enriched genes in human (Fig. 2). For example, *NPY* is enriched in subplate at 21 pcw (but not 15 or 16 pcw) as shown both by microarray and *in situ* hybridization (Fig. 4a). To facilitate a comparative analysis, we identified a high confidence set of 150 subplate-enriched genes in human or mouse (Supplementary Table 8). Many genes showed similar enrichment in the developing mouse and human subplate, including the known subplate markers *KCNAB1* and *NR4A2* (refs 17, 18 and 44) (Fig. 4b). Several genes showed enrichment in developing human but not mouse subplate, including the hypocretin (orexin)

Figure 4 | Common and distinct subplate markers in human and mouse. a, *NPY* is enriched in subplate at 21 pcw but not 15 or 16 pcw based on microarray (left) and *in situ* hybridization (right). Microarray data are plotted as the average ± standard error of the mean (s.e.m.) for each layer in each brain. **b**, Genes with subplate enrichment in both species, based on microarray data in human (top) and *in situ* hybridization data at embryonic (E) day E15.5 or E18.5 mouse (bottom). **c**, Genes with subplate enrichment in human but either no expression (*DKK1*, *HCRTR2* and *MESP1*) or no subplate enrichment (*CDH1* and *CRYM*) in mouse. **d**, Genes with subplate enrichment in developing mouse, but not human subplate. Asterisks indicate common expression between human and mouse in other layers (that is, subpial granular zone and pia mater). Mouse *in situ* hybridization images taken from the Allen Developing Mouse Brain Atlas.
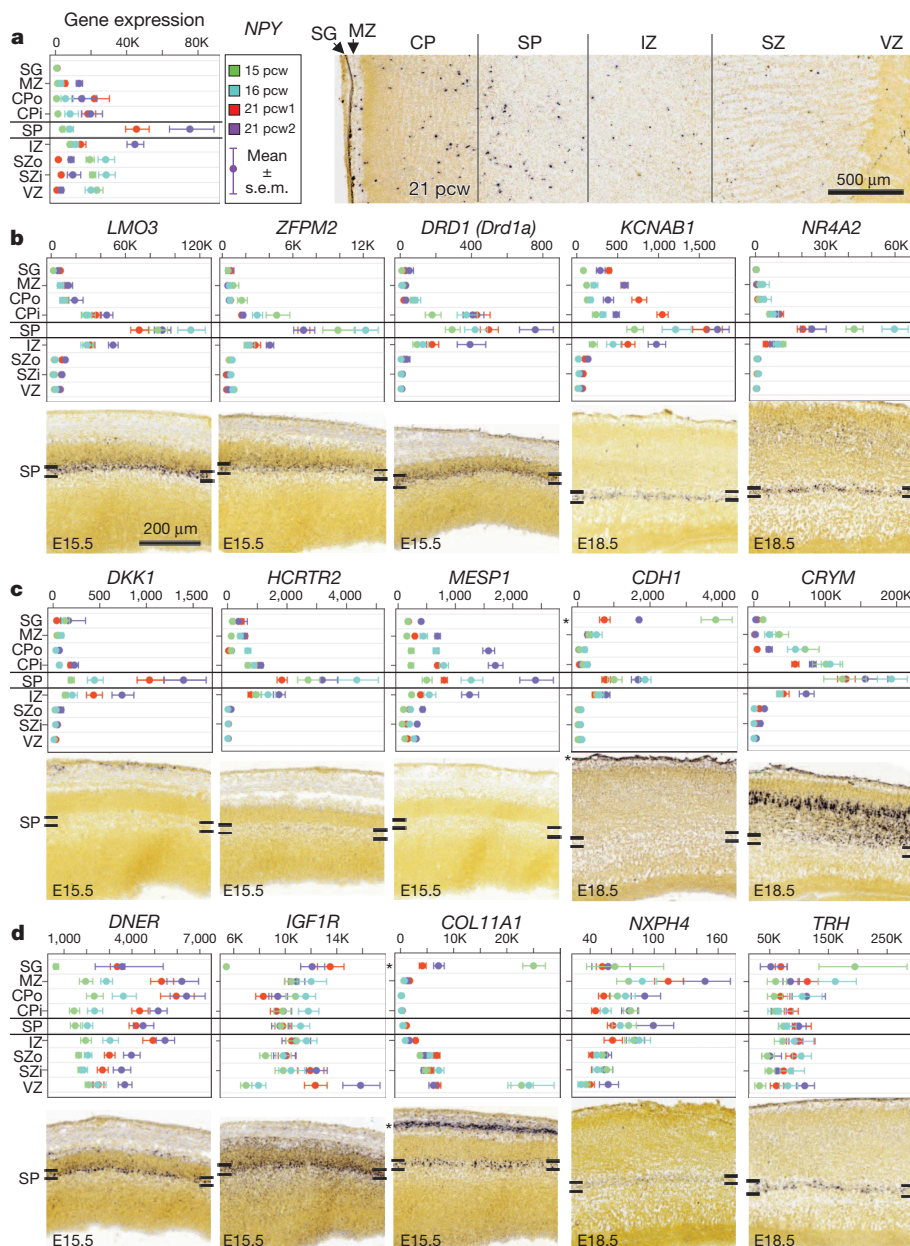
receptor *HCRTR2* (Fig. 4c and Extended Data Fig. 8a), which is thought to regulate sleep-wakefulness and is highly expressed in mouse hypothalamus[45]. Conversely, *Trh* and *Nxph4* show enriched expression in mouse but not human subplate (Fig. 4d and Extended Data Fig. 8b). *Trh* is also not expressed in the rat subplate[18], suggesting this pattern is specific to mouse. These results indicate that the evolutionary elaboration of subplate in primates is associated with altered gene expression.

Developmental gradients in neocortex

Cortical patterning is likely a result of intrinsic signalling, controlled in part by graded expression of transcription factors during early cortical development, followed by extrinsic signalling from thalamic afferents after the start of corticogenesis[46–49]. We sought to identify putative patterning centres, defined as regions where many genes show peak expression tapering off with distance, for each layer of the human prenatal cortex using an unbiased approach. To do so, we first assigned three-dimensional (3D) coordinates to each cortical sample, and then identified the location of maximum expression of the most graded genes in all four brains. In several layers, including CPo and SZi (Fig. 5a, b), most of these genes peaked in the frontal or temporal lobes. Rather than peaks in presumptive functional areas, this suggests a generally rostro-caudal organization axis that is better characterized as fronto-temporal, following the contour of the developing cortex. To identify such fronto-temporal patterning genes directly, we correlated gene expression in each cortical layer against the angular position of each neocortical region, as illustrated schematically in Fig. 5c. All layers contained gradient genes conserved across all four brains ($P \sim 0$, permutation analysis; FDR < 2.4%; Supplementary Methods), and each layer contained distinct sets of gradient genes, particularly when comparing germinal with post-mitotic layers (Supplementary Table 9). Gradient genes in the ventricular zone probably reflect intrinsic areal specification as the ventricular zone does not receive thalamic innervation. Our results indicate that rostral and caudal genes that were identified in grossly dissected cortex[3] mainly represent gradients in post-mitotic cells, as they show significant overlap with gradient genes in marginal zone, cortical plate and intermediate zone (Extended Data Fig. 9).

Some features of areal patterning appear to be preserved between mouse and human. For example, *FGFR3*, which is known to cause defects in human temporal cortex when mutated[50], shows significant caudal enrichment in all germinal layers in both species (Fig. 5d, e). Conversely, *CBLN2* shows rostral enrichment in human CPo (Fig. 5f) and in mouse (Fig. 5g), with an abrupt expression cutoff, implicating *CBLN2* in early rostral patterning. Comparing these data to a microarray analysis of rostral versus caudal cortex in E14 mouse[51] identified 20 additional genes with consistent rostrocaudal gradation between species (Supplementary Table 9 and Supplementary Methods). We find more rostrally than caudally enriched genes in nearly every cortical layer (Fig. 5h), whereas two studies of gradient expression in prenatal mouse brain identified more[49] or comparable[51] caudally enriched genes, indicating potential species differences related to areal patterning. This human frontal asymmetry is most apparent in the outer post-mitotic layers, the marginal zone and CPo (Fig. 5i) and in SZo, which generates most superficial cortical plate neurons in primate[7,8], indicating that these genes may play a part in the expansion and reorganization of human prefrontal cortex[8]. Alternatively, this asymmetry could reflect temporal differences, as peak generation of excitatory neurons in visual cortex is delayed relative to frontal cortex[52].

Patterned expression of genes near haCNSs

Conserved non-coding sequences (CNSs) are genomic regions with exceptionally high similarity across divergent organisms, and therefore thought to be important for organism viability. CNSs are typically located by genes important for developmental regulation, and many show significant enhancer activity in brain[5]. Genes near CNSs with significantly accelerated rates of substitution in the human lineage (haCNSs)[15] are particularly likely to show differential expression between regions of developing human neocortex[3,5], indicating transcriptional regulation by haCNSs may be important in human-specific neurodevelopment.
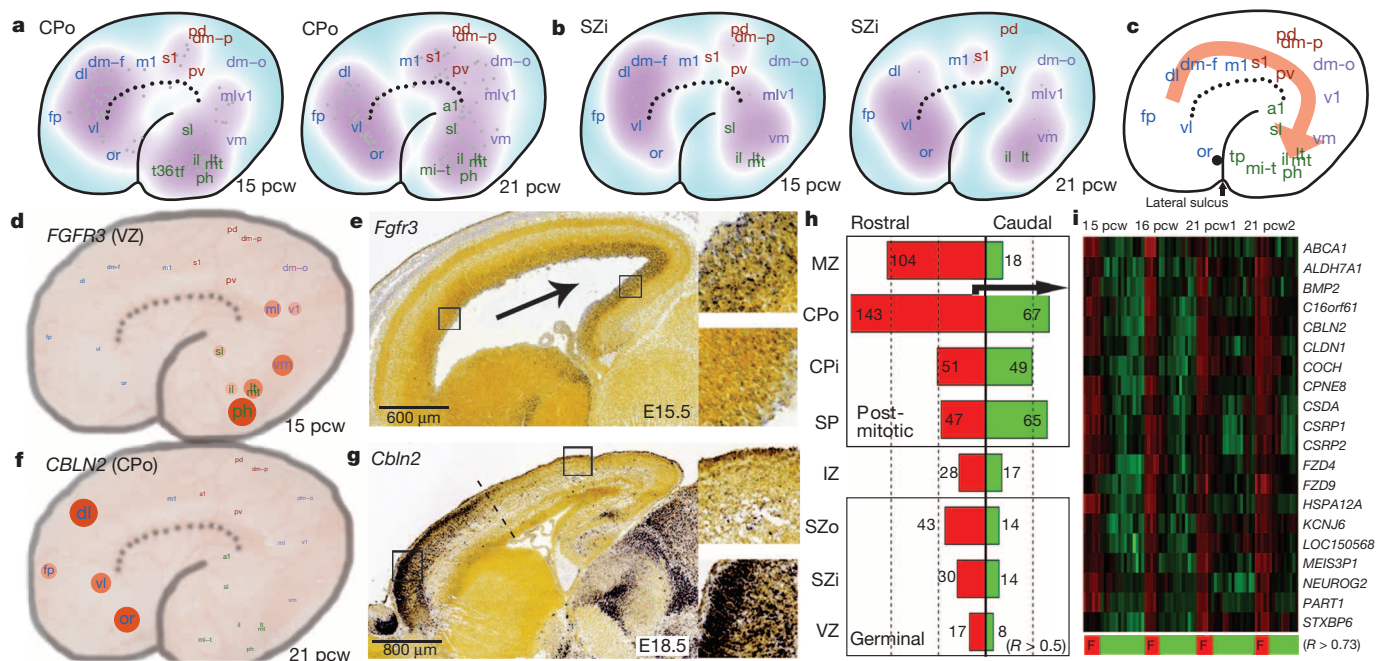
Figure 5 | Areal patterning in the developing neocortex. a, b, Density plot showing the location of highest expression (colour coded purple) for genes with gradient-like expression in CPo (**a**) or SZi (**b**) in each brain. **c,** Schematic illustrating the predominant direction of gene gradients, which follow a fronto-temporal trajectory. **d,** *FGFR3* shows caudal enrichment in germinal zones of the developing human cortex. Samples are plotted on a schematic of the prenatal cortex, with expression level indicated by circle size and colour. **e,** Similar enrichment is seen in the developing mouse cortex. **f, g,** *CBLN2* shows rostral enrichment in cortical plate of both human (**f**) and mouse (**g**). **h,** Barplot showing common rostrally (red) and caudally (green) enriched genes across brains for each layer in humans. **i,** Heatmap representation of the top 20 rostrally enriched genes in the CPo shows selective enrichment in frontal (F) lobe (F, in bottom colour bar).

Table 1 | Human-specific neurodevelopmental processes enriched for genes near haCNSs

Category	Subcategory	O	O/E	Total (%)	P value
Areal	All	46	3.54	9.9	2.56×10^{-14}
	Rostral (R)	34	4.24	11.8	2.60×10^{-13}
	Caudal (C)	14	2.77	7.7	1.78×10^{-4}
	Post-mitotic (P)	35	3.37	9.4	1.01×10^{-14}
	IZ	11	8.80	24.4	2.34×10^{-9}
	Germinal (G)	7	2.68	7.4	4.65×10^{-3}
Network	C31 (interneuron)	18	5.31	14.8	1.15×10^{-9}
	C16 (subplate)	25	3.98	11.1	9.80×10^{-10}

O, observed; E, expected. Areal genes include all genes in Fig. 5h. No other layers (Fig. 2) or modules (Fig. 3) were significantly enriched for genes near haCNSs.

Our results confirm and extend these findings (Table 1 and Supplementary Table 9). Rostrally enriched genes include significantly more haCNSs than caudally enriched genes, consistent with the expanded frontal cortex in primates and the developmental role of haCNSs. We also find more haCNSs with areal expression patterns in post-mitotic than germinal layers, even after accounting for the larger number of areal genes in post-mitotic layers. Surprisingly, nearly 25% of regional genes in the intermediate zone (11/45) were haCNSs. This result, which cannot be explained by over-representation of neural adhesion genes in the intermediate zone, suggests that the intermediate zone is of particular importance in areal cortical identity during human development[53].

Finally, we assessed the distribution of haCNSs in modules from our whole cortex network (Fig. 3a–c). Only two of the 42 tested modules were enriched for haCNSs: the interneuron-related (C31) and subplate-enriched (C16) modules, which both mark processes with features potentially distinct to the human lineage. Interestingly, *FOXP2*, implicated in the specialization of language areas[54], is included in module C16, and shows enrichment in parietal and temporal lobes including presumptive language areas (Extended Data Fig. 10). No layers other than subplate were significantly enriched for haCNSs (data not shown). Together these results support the hypothesis that transcriptional networks underlying the evolution of human neocortex can at least partly be traced to haCNSs.

Discussion

Studies of the developing human brain are essential for elucidating the details of human brain formation, function and evolutionary differences, and for understanding developmental mechanisms underlying neurodevelopmental disorders such as autism[39,40] and schizophrenia[19]. The atlas of the mid-gestational human brain described here, part of the BrainSpan Atlas of the Developing Human Brain, builds on digital molecular brain atlasing efforts in mouse[20,22,24] and adult human[23] by providing transcriptome resources on prenatal specimens typically inaccessible for research. Several recent studies have assayed a limited set of brain structures[1,3–5] and layers[2] from prenatal human brain. In contrast, the current project aimed for anatomical comprehensiveness at a fine nuclear/laminar level, albeit with a small number of specimens. This degree of specificity necessitated using available methods for small sample amplification and DNA microarrays (the same platform recently used for adult human[23]), but newer techniques may soon allow moving to the resolution of single cells using RNA sequencing for complete transcriptome coverage[55].

Many differences in cortical development between human, non-human primate and rodent have been documented, including an expanded subplate zone[8] and subpial granular zone[9], expansion of association areas particularly in frontal lobe[8], expansion of superficial layers that greatly increase the extent of cortico-cortical connections, and the appearance of a secondary proliferative zone, the SZo, that probably allow the massive expansion of human cortex[6,7]. We find transcriptional features related to each of these anatomical features, although we were able to identify only minimal molecular differences between the inner and outer subventricular zones, leaving open the question of what distinguishes this

primate-specific zone of cortical precursors. These data also provide a powerful map to pin an anatomical and developmental locus on genes related to neurodevelopmental disease origins and human-specific brain function and evolution. Although the current analyses only scratch the surface, these data will be extremely useful for generating and testing new hypotheses about molecular substrates for specific features of human brain development and function.

METHODS SUMMARY

Anatomically comprehensive transcriptional profiling of prenatal human brains used high-throughput tissue processing and data generation pipelines for post-mortem brain anatomical delineation, sample isolation and microarray analysis. Data visualization and mining tools were developed to create a publicly accessible data resource (http://www.brainspan.org/).

Post-mortem tissue acquisition and screening. Tissue was provided by the Birth Defects Research Laboratory (BDRL) at the University of Washington and Advanced Bioscience Resources (ABR; Alameda, California). Tissue with no known history of maternal drug or alcohol abuse, potential teratogenic events, or HIV1/2, HepB or HepC infection, was eligible for inclusion in the study. Due to confidentiality standards of tissue providers, it is impossible to rule out potential neuropathologies in this tissue; however, no neuropathological defects were observed in histological data derived from these tissues. Eligible tissue was screened to ensure cytoarchitectural integrity (analysis of Nissl-stained sections) and high RNA quality. RNA integrity numbers (RINs) ranged from 8.2–9. All work was performed following guidelines for the research use of human brain tissue (ABR) or the UAGA and NOTA guidelines for acquisition and distribution of human tissue for bio-medical research purposes (BDRL) and with approval by the Human Investigation Committees and Institutional Ethics Committees of each institute. Appropriate written informed consent was obtained and all available non-identifying information was recorded for each sample. Four specimens were used for microarray profiling (Supplementary Table 1).

Sample processing. For each specimen, the frozen left hemisphere was cut into ~2.0-cm-thick slabs. Slabs were sectioned to allow histological staining, as shown in Fig. 1d. Total RNA was isolated and microarray data were generated by Covance Genomics Laboratory (Seattle, Washington) on Agilent $8 \times 60K$ custom design arrays (AMADID no. 024915)[20].

Microarray data analysis. Data were preprocessed by removing outliers and then normalizing. Differential expression was assessed using correlation to a template vector or by measuring fold change. WGCNA was performed as previously described[36]. Lists were characterized using gene ontology and cell type enrichments. Enrichment of haCNSs was quantified using hypergeometric tests.

***In situ* hybridization.** *In situ* hybridization used a semi-automated non-isotopic technology platform as previously described[24], with some modifications such as a reduction in proteinase K concentration. Riboprobes were designed to overlap probe designs for homologous mouse genes in the Allen Developing Mouse Brain Atlas (http://developingmouse.brain-map.org/). FISH was performed as previously described[56].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478,** 519–523 (2011).
2. Fietz, S. A. *et al.* Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proc. Natl Acad. Sci. USA* **109,** 11836–11841 (2012).
3. Johnson, M. B. *et al.* Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62,** 494–509 (2009).
4. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478,** 483–489 (2011).
5. Lambert, N. *et al.* Genes expressed in specific areas of the human fetal cerebral cortex display distinct patterns of evolution. *PLoS ONE* **6,** e17753 (2011).
6. Kriegstein, A., Noctor, S. & Martínez-Cerdeño, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. *Nature Rev. Neurosci.* **7,** 883–890 (2006).
7. Hansen, D. V., Lui, J., Parker, P. & Kriegstein, A. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464,** 554–561 (2010).
8. Smart, I., Dehay, C., Giroud, P., Berland, M. & Kennedy, H. Unique morphological features of the proliferative zones and postmitotic compartments of the neural epithelium giving rise to striate and extrastriate cortex in the monkey. *Cereb. Cortex* **12,** 37–53 (2002).

9. Brun, A. The subpial granular layer of the foetal cerebral cortex in man. Its ontogeny and significance in congenital cortical malformations. *Acta Pathologica et Microbiologica Scandinavica* **179** (suppl.), 3–98 (1965).

10. Anderson, S. A., Eisenstat, D. D., Shi, L. & Rubenstein, J. L. Interneuron migration from basal forebrain to neocortex: dependence on *Dlx* genes. *Science* **278**, 474–476 (1997).

11. Ma, T. *et al.* Subcortical origins of human and monkey neocortical interneurons. *Nature Neurosci.* **16**, 1588–1597 (2013).

12. Yu, X. & Zecevic, N. Dorsal radial glial cells have the potential to generate cortical interneurons in human but not in mouse brain. *J. Neurosci.* **31**, 2413–2420 (2011).

13. Letinic, K., Zoncu, R. & Rakic, P. Origin of GABAergic neurons in the human neocortex. *Nature* **417**, 645–649 (2002).

14. Hansen, D. V. *et al.* Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nature Neurosci.* **16**, 1576–1587 (2013).

15. Prabhakar, S., Noonan, J., Pääbo, S. & Rubin, E. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**, 786 (2006).

16. Ayoub, A. E. *et al.* Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing. *Proc. Natl Acad. Sci. USA* **108**, 14950–14955 (2011).

17. Montiel, J. *et al.* Hypothesis on the dual origin of the mammalian subplate. *Front. Neuroanat.* **5**, 25 (2011).

18. Wang, W. Z. *et al.* Comparative aspects of subplate zone studied with gene expression in sauropsids and mammals. *Cereb. Cortex* **21**, 2187–2203 (2011).

19. Insel, T. R. Rethinking schizophrenia. *Nature* **468**, 187–193 (2010).

20. Paxinos, G. & Franklin, K. *Paxinos and Franklin's the Mouse Brain in Stereotaxic Coordinates* 4th edn (Academic Press, 2012).

21. Saleem, K. & Logothetis, N. *A Combined MRI and Histology Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates* 2nd edition (Academic Press, 2012).

22. Dong, H. W. *Allen Reference Atlas: a Digital Color Brain Atlas of the C57BL/6J Male Mouse* (Wiley, 2008).

23. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).

24. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

25. Bernard, A. *et al.* Transcriptional architecture of the primate neocortex. *Neuron* **73**, 1083–1099 (2012).

26. Steinfeld, R. *et al.* Folate receptor alpha defect causes cerebral folate transport deficiency: a treatable neurodegenerative disorder associated with disturbed myelin metabolism. *Am. J. Hum. Genet.* **85**, 354–363 (2009).

27. Dubourg, C. *et al.* Molecular screening of *SHH, ZIC2, SIX3*, and *TGIF* genes in patients with features of holoprosencephaly spectrum: mutation review and genotype–phenotype correlations. *Hum. Mutat.* **24**, 43–51 (2004).

28. Huang, H. *et al.* Anatomical characterization of human fetal brain development with diffusion tensor magnetic resonance imaging. *J. Neurosci.* **29**, 4263–4273 (2009).

29. Bystron, I., Blakemore, C. & Rakic, P. Development of the human cerebral cortex: Boulder Committee revisited. *Nature Rev. Neurosci.* **9**, 110–122 (2008).

30. Vogt Weisenhorn, D. M., Prieto, E. W. & Celio, M. R. Localization of calretinin in cells of layer I (Cajal-Retzius cells) of the developing cortex of the rat. *Brain Res. Dev. Brain Res.* **82**, 293–297 (1994).

31. Englund, C. *et al.* Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. *J. Neurosci.* **25**, 247–251 (2005).

32. Levitt, P. & Rakic, P. Immunoperoxidase localization of glial fibrillary acidic protein in radial glial cells and astrocytes of the developing rhesus monkey brain. *J. Comp. Neurol.* **193**, 815–840 (1980).

33. Grinberg, I. *et al.* Heterozygous deletion of the linked genes *ZIC1* and *ZIC4* is involved in Dandy-Walker malformation. *Nature Genet.* **36**, 1053–1055 (2004).

34. Inoue, T., Ogawa, M., Mikoshiba, K. & Aruga, J. *Zic* deficiency in the cortical marginal zone and meninges results in cortical lamination defects resembling those in type II lissencephaly. *J. Neurosci.* **28**, 4712–4725 (2008).

35. Kostović, I. & Judas, M. The development of the subplate and thalamocortical connections in the human foetal brain. *Acta Paediatr.* **99**, 1119–1127 (2010).

36. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17 (2005).

37. Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nature Neurosci.* **11**, 1271–1282 (2008).

38. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).

39. Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).

40. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).

41. Gaspard, N. & Vanderhaeghen, P. Laminar fate specification in the cerebral cortex. *F1000 Biol. Rep.* **3**, 6 (2011).

42. Betizeau, M. *et al.* Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. *Neuron* **80**, 442–457 (2013).

43. Kawaguchi, A. *et al.* Single-cell gene profiling defines differential progenitor subclasses in mammalian neurogenesis. *Development* **135**, 3113–3124 (2008).

44. Butler, D. M., Ono, J. K., Chang, T., McCaman, R. E. & Barish, M. E. Mouse brain potassium channel β1 subunit mRNA: cloning and distribution during development. *J. Neurobiol.* **34**, 135–150 (1998).

45. Gerashchenko, D. *et al.* Hypocretin-2-saporin lesions of the lateral hypothalamus produce narcoleptic-like sleep behavior in the rat. *J. Neurosci.* **21**, 7273–7283 (2001).

46. Mallamaci, A. & Stoykova, A. Gene networks controlling early cerebral cortex arealization. *Eur. J. Neurosci.* **23**, 847–856 (2006).

47. O'Leary, D. D. Do cortical areas emerge from a protocortex? *Trends Neurosci.* **12**, 400–406 (1989).

48. Rakic, P. Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).

49. Sansom, S. N. & Livesey, F. Gradients in the brain: the control of the development of form and function in the cerebral cortex. *Cold Spring Harb. Perspect. Biol.* **1**, a002519 (2009).

50. Hevner, R. F. The cerebral cortex malformation in thanatophoric dysplasia: neuropathology and pathogenesis. *Acta Neuropathol.* **110**, 208–221 (2005).

51. Pinto, L. *et al.* AP2γ regulates basal progenitor fate in a region- and layer-specific manner in the developing cortex. *Nature Neurosci.* **12**, 1229–1237 (2009).

52. Rakic, P. Neurogenesis in adult primates. *Prog. Brain Res.* **138**, 3–14 (2002).

53. Allendoerfer, K. L. & Shatz, C. J. The subplate, a transient neocortical structure: its role in the development of connections between thalamus and cortex. *Annu. Rev. Neurosci.* **17**, 185–218 (1994).

54. Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).

55. Qiu, S. *et al.* Single-neuron RNA-seq: technical feasibility and reproducibility. *Front. Genet.* **3**, 124 (2012).

56. Sorensen, S. A. *et al.* Correlated gene expression and target specificity demonstrate excitatory projection neuron diversity. *Cereb. Cortex* http://dx.doi.org/10.1093/cercor/bht243 (2013).

Author Contributions E.S.L, S.-L.D., K.A.S. and S.M.S. contributed significantly to the overall project design. S.M.S, K.A.S., A.E., A.B., and P.W. managed the tissue and sample processing in the laboratory. K.A., J.M.A., C.B., D.B., K.B., S.B., S.C., A.C., C.C., R.A.D., G.Ge., J.G., L.G., B.W.G., R.E.H., T.A.L., Na.M., N.F.M., N.-K.N., A.O., E.O., J.Pa., P.D.P., S.E.P., M.P., Me.R., J.J.R., K.R., D.S., Me.S., S.S., N.V.S. and Mi.S. contributed to tissue and sample processing. E.H.S., Z.L.R., T.N.-C., and I.A.G. contributed to establishing the tissue acquisition pipeline. N.D., J.N. and A.B. contributed to protocol development. A.S.P., L.Z., B.F., and H.H. contributed to MR and DWI imaging and analysis. J.M.J., C.R.S., and D.W. provided engineering support. S.-L.D., R.A.D., P.D.P., D.S. and J.G.H. contributed to the neuroanatomical design and implementation. S.-L.D., B.A.C.F., Ph.L., B.M., J.J.R., R.F.H., N.Se. and J.G.H. contributed to the reference atlas design, quality control and implementation. L.N., A.S. and C.D. managed the creation of the data pipeline, visualization and mining tools. L.N., A.S., T.A.D., D.F., T.P.F., G.Gu, C.L.K., C.La., F.L., N.Sj. and A.J.S. contributed to the creation of the data pipeline, visualization and mining tools. J.A.M., S.-L.D., R.F.H., C.-K.L., M.J.H., S.M.S. and E.S.L. contributed to data analysis and interpretation. M.B.G., D.H.G., J.A.K., Pa.L., J.W.P., N.Se. and A.R.J. contributed to overall project design and consortium management. E.S.L. and M.J.H. conceived the project, and the manuscript was written by J.A.M. and E.S.L. with input from all other authors.

Author Information These data are freely accessible as part of the BrainSpan Atlas of the Developing Human Brain (http://brainspan.org), also available via the Allen Brain Atlas data portal (http://www.brain-map.org). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.L. (edl@alleninstitute.org).

ARTICLE

A mesoscale connectome of the mouse brain

Seung Wook Oh[1]*, Julie A. Harris[1]*, Lydia Ng[1]*, Brent Winslow[1], Nicholas Cain[1], Stefan Mihalas[1], Quanxin Wang[1], Chris Lau[1], Leonard Kuan[1], Alex M. Henry[1], Marty T. Mortrud[1], Benjamin Ouellette[1], Thuc Nghi Nguyen[1], Staci A. Sorensen[1], Clifford R. Slaughterbeck[1], Wayne Wakeman[1], Yang Li[1], David Feng[1], Anh Ho[1], Eric Nicholas[1], Karla E. Hirokawa[1], Phillip Bohn[1], Kevin M. Joines[1], Hanchuan Peng[1], Michael J. Hawrylycz[1], John W. Phillips[1], John G. Hohmann[1], Paul Wohnoutka[1], Charles R. Gerfen[2], Christof Koch[1], Amy Bernard[1], Chinh Dang[1], Allan R. Jones[1] & Hongkui Zeng[1]

Comprehensive knowledge of the brain's wiring diagram is fundamental for understanding how the nervous system processes information at both local and global scales. However, with the singular exception of the *C. elegans* microscale connectome, there are no complete connectivity data sets in other species. Here we report a brain-wide, cellular-level, mesoscale connectome for the mouse. The Allen Mouse Brain Connectivity Atlas uses enhanced green fluorescent protein (EGFP)-expressing adeno-associated viral vectors to trace axonal projections from defined regions and cell types, and high-throughput serial two-photon tomography to image the EGFP-labelled axons throughout the brain. This systematic and standardized approach allows spatial registration of individual experiments into a common three dimensional (3D) reference space, resulting in a whole-brain connectivity matrix. A computational model yields insights into connectional strength distribution, symmetry and other network properties. Virtual tractography illustrates 3D topography among interconnected regions. Cortico-thalamic pathway analysis demonstrates segregation and integration of parallel pathways. The Allen Mouse Brain Connectivity Atlas is a freely available, foundational resource for structural and functional investigations into the neural circuits that support behavioural and cognitive processes in health and disease.

A central principle of neuroscience is that the nervous system is a network of diverse types of neurons and supporting cells communicating with each other mainly through synaptic connections. This overall brain architecture is thought to be composed of four systems—motor, sensory, behavioural state and cognitive—with parallel, distributed and/or hierarchical sub-networks within each system and similarly complex, integrative interconnections between different systems[1]. Specific groups of neurons with diverse anatomical and physiological properties populate each node of these sub- and supra-networks, and form extraordinarily intricate connections with other neurons located near and far. Neuronal connectivity forms the structural foundation underlying neural function, and bridges genotypes and behavioural phenotypes[2,3]. Connectivity patterns also reflect the evolutionary conservation and divergence in brain organization and function across species, as well as both the commonality among individuals within a given species and the uniqueness of each individual brain.

Despite the fundamental importance of neuronal connectivity, our knowledge of it remains remarkably incomplete. *C. elegans* is the only species for which an essentially complete wiring diagram of its 302 neurons has been obtained through electron microscopy[4]. Histological tract tracing studies in a wide range of animal species has generated a rich body of knowledge that forms the foundation of our current understanding of brain architecture, such as the powerful idea of multi-hierarchical processing in sensory cortical systems[5]. However, much of these data are qualitative, incomplete, variable, scattered and difficult to retrieve. Thus, our knowledge of whole-brain connectivity is fragmented, without a cohesive and comprehensive understanding in any single vertebrate animal species (see for example the BAMS database for the rat brain[6]). With recent advances in both computing power and optical imaging techniques, it is now feasible to systematically map connectivity throughout the entire brain. A salient example of this is the ongoing effort in mapping connections in the *Drosophila* brain[7,8].

The connectome[9] refers to a comprehensive description of neuronal connections, for example, the wiring diagram of the entire brain. Given the enormous range of connectivity in the mammalian brain and the relative inaccessibility of the human brain, such descriptions can exist at multiple levels: macro-, meso- or microscale. At the macroscale, long-range, region-to-region connections can be inferred from imaging white-matter fibre tracts through diffusion tensor imaging (DTI) in the living brain[10]. However, this is far from cellular-level resolution, given the size of single volume elements (voxels >1 mm[3]). At the microscale, connectivity is described at the level of individual synapses, for example, through electron microscopic reconstruction at the nanometer scale[4,11–15]. At present, the enormous time and resources required for this approach makes it best suited for relatively small volumes of tissue (<1 mm[3]). At the mesoscale, both long-range and local connections can be described using a sampling approach with various neuroanatomical tracers that enable whole-brain mapping in a reasonable time frame across many animals. In addition, cell-type-specific mesoscale projects have the potential to dramatically enhance our understanding of the brain's organization and function because cell types are fundamental cellular units often conserved across species[16,17].

Here we present a mesoscale connectome of the adult mouse brain, The Allen Mouse Brain Connectivity Atlas. Axonal projections from regions throughout the brain are mapped into a common 3D space using a standardized platform to generate a comprehensive and quantitative database of inter-areal and cell-type-specific projections. This Connectivity Atlas has all the desired features summarized in a mesoscale connectome position essay[18]: brain-wide coverage, validated and versatile experimental techniques, a single standardized data format, a quantifiable

[1]Allen Institute for Brain Science, Seattle, Washington 98103, USA. [2]Laboratory of Systems Neuroscience, National Institute of Mental Health, Bethesda, Maryland 20892, USA.
*These authors contributed equally to this work.

and integrated neuroinformatics resource and an open-access public online database.

Creating the Allen Mouse Brain Connectivity Atlas

A standardized data generation and processing platform was established (Fig. 1a, see Methods). Recombinant adeno-associated virus (AAV), serotype 1, expressing EGFP optimally was chosen as the anterograde tracer to map axonal projections[19,20]. We also confirmed that AAV was at least as efficient as, and more specific than, the classic anterograde tracer biotinylated dextran amine (BDA) (Extended Data Fig. 1), as described separately[21].

EGFP-labelled axonal projections were systematically imaged using the TissueCyte 1000 serial two-photon (STP) tomography system[22], which couples high-speed two-photon microscopy with automated vibratome sectioning of an entire mouse brain. High x–y resolution (0.35 µm) 2D images in the coronal plane were obtained at a z-sampling interval of 100-µm across the entire brain during a continuous 18.5 h scanning period, resulting in 140 serial sections (a ~750 gigabyte (GB) data set) for each brain (Extended Data Fig. 2a and Supplementary Video 1). Owing to its block-face imaging nature, STP tomography essentially eliminates tissue distortions that occur in conventional section-based histological methods and provides a series of highly conformed, inherently pre-aligned images amenable to precise 3D mapping.

Image series were processed in an informatics pipeline with a series of modules (see Methods). The injection site location of each brain was manually drawn and annotated using the Allen Reference Atlas[23] and other reference data sets when appropriate. Stringent quality control criteria were applied, discarding ~25% of all scanned brains due to insufficient quality in labelling or imaging. Each image set was registered into a 3D Allen Reference Atlas model in two steps (Fig. 1b, upper panels). First, a registration template was created by averaging many image sets, and every image stack was aligned to this average template brain. This process was repeated for multiple rounds, first globally (affine registration) and then locally (deformable registration), each round generating a better average template and more precise alignment of individual brains. The final average template brain, averaged from 1,231

brains, shows remarkably clear anatomical features and boundaries. Second, the average template brain was aligned with the 3D reference model, again using local alignment (Supplementary Video 2).

We developed a signal detection approach and applied it to each section to segment GFP signals from background (Fig. 1b, lower panels). Signals within injection site polygons were computed separately from the rest of the brain. The segmented pixel counts were gridded into $100 \times 100 \times 100$ µm^3 voxels to create an isotropic 3D summary of the projection data. These voxels were used for data analysis, real-time data and correlative searches, and visualization of projection relationships in the Brain Explorer.

Meaningful informatics data quantification and comparison relies on the mapping precision of the raw data sets into the 3D reference framework. We investigated registration variability in two ways. First, we selected 10 widely distributed anatomical fiducial points to compare variability among 30 randomly selected brains (Extended Data Fig. 2b). We found a high degree of concordance among individual brains, with median variation < 49 µm in each dimension between each brain and the average template brain, which is comparable to the median inter-rater variation of < 39 µm. The median difference is < 71 µm between each brain and the Reference Atlas. Second, we compared manual and informatics annotations of the injection sites from all Phase I (see below) brains. The informatics-derived assignment of injection site structures had $> 75\%$ voxel-level concordance with manual expert annotation for almost all injection sites (Extended Data Fig. 2c). These analyses confirmed the relatively high fidelity of co-registration of raw image data with the Allen Reference Atlas. The remaining difference is mainly due to the imperfect alignment between the average template brain and the Nissl-section-based Reference Atlas (Supplementary Video 2).

Mapping axonal projections in the whole mouse brain

The connectivity mapping was carried out in two phases. In Phase I (regional projection mapping), axonal projections from 295 non-overlapping anatomical regions, defined from the Allen Reference Atlas ontology and tiling the entire brain space (Supplementary Table 1), were characterized in wild-type mice with a pan-neuronal AAV vector expressing
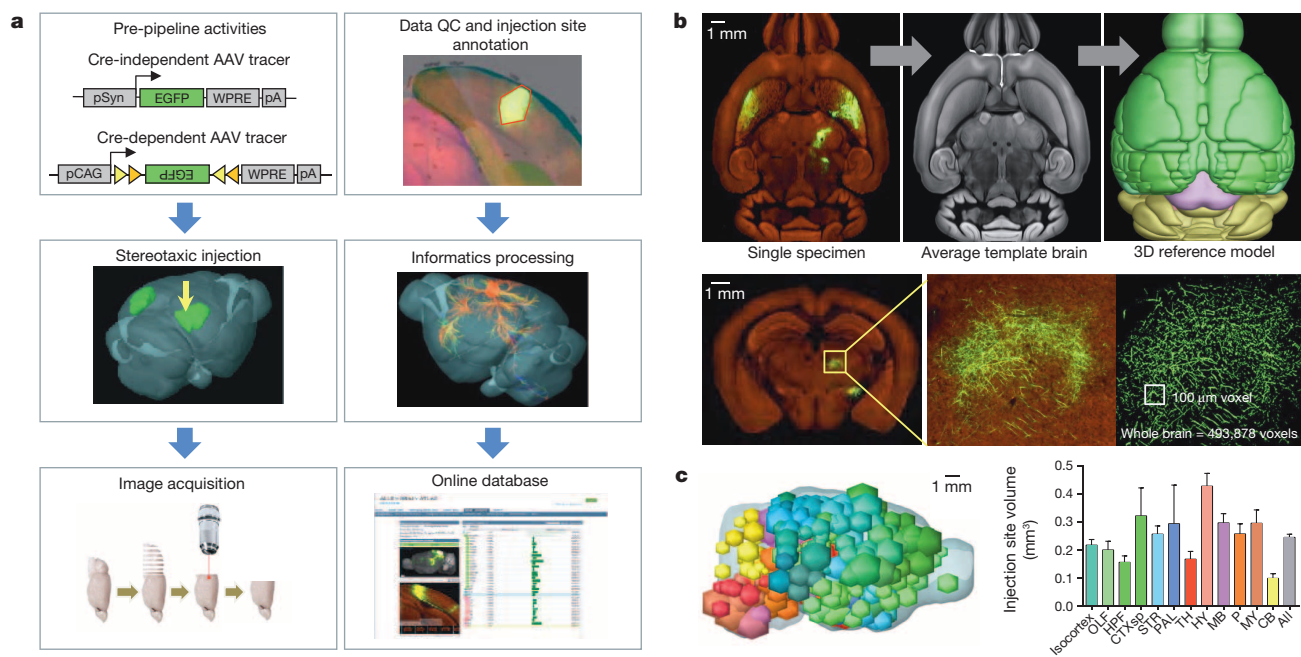
Figure 1 | Creation of the Connectivity Atlas. a, The data generation and processing pipeline. QC, quality control. **b**, The two main steps of informatics data processing: registration of each image series to a 3D template (upper panels) and segmentation of fluorescent signal from background (lower panels). **c**, Distribution of injection sites across the brain. The volume of the

injection was calculated and represented as a sphere. Locations of all these injection spheres are superimposed together (left panel). Mean injection volumes ($\pm$ s.e.m.) across major brain subdivisions are shown (right panel, see Extended Data Fig. 3).

EGFP under the human synapsin I promoter (AAV2/1.pSynI.EGFP.WPRE.bGH, Fig. 1a). In Phase II (Cre driver based projection mapping), axonal projections from genetically defined neuronal populations are characterized in Cre driver mouse lines with a Cre-dependent AAV (AAV2/1.pCAG.FLEX.EGFP.WPRE.bGH, Fig. 1a). We only report here on the completed Phase I study, which includes 469 image sets with injection sites covering nearly the entire brain (Fig. 1c, Extended Data Fig. 3 and Supplementary Video 3). Only 18 intended structures were completely missed due to redundancy or injection difficulty (Supplementary Table 1).

We examined multiple projection data sets in detail and found that they were complete in capturing all known projection target sites throughout the brain, sensitive in detecting thin axon fibres, and consistent in quality to allow qualitative and quantitative comparisons. As an example, 7 representative isocortical injections (Fig. 2) reveal distinct projection patterns in the striatum, thalamus, zona incerta, midbrain, pons and medulla. To compare the brain-wide spatial distribution of projections between cortical source regions, we placed each isocortical injection experiment into one of 9 broad functional groups: frontal, motor, anterior cingulate, somatosensory, auditory, retrosplenial, visual, ventral and associational areas (Extended Data Fig. 4). The average percentages of total projection signals into 12 major brain subdivisions showed disproportionately large projections within the isocortex, as well as distinct subcortical distributions.

Brain–wide connectivity matrix

After segmentation and registration, we derived quantitative values from segmented signals in each of the ~500,000 voxels contained within each brain. We constructed a brain-wide, inter-areal, weighted connectivity matrix using the entire Phase I experimental data set (Fig. 3, see Supplementary Table 2 for the underlying values). The Allen Reference Atlas contains 863 grey-matter structures at the highest level of the ontology tree (Supplementary Table 1). We focused our analyses on the chosen 295 structures, which are at a mid-ontology level corresponding best with the approximate size of the tracer infection areas (for example, isocortical areas are not subdivided by layers in this matrix),

but our techniques may be used at deeper levels in future studies. The projection signal strength between each source and target was defined as the total volume of segmented pixels in the target (summed across all voxels within each target), normalized by the injection site volume (total segmented pixels within the manually drawn injection area).

The majority of the 469 Phase I image sets are single injections into spatially distinct regions, but a subset of these are repeated injections into the same regions. To assess the consistency of projection patterns across different animals and the reliability of using a single experiment to define connections from any particular region, we compared brain-wide connectivity strengths in 12 sets of duplicate injections (Extended Data Fig. 5). Each pair was highly correlated across a range of projection strengths. Differences between any two points were on average only a half order of magnitude (within one standard deviation). In primate cortex, single tracer injections were also found to reliably predict mean values obtained from repeated injections into the same source[24].

The AAV tracer expresses cytoplasmic EGFP, which labels all processes of the infected neuron, including axons and synaptic terminals. Signals associated with the major fibre tracts of the brain, marked in the Allen Reference Atlas, were removed before the informatics quantification. However, there are also areas (for example, striatum) where axons pass through without making synapses. Although passing fibres can generally be distinguished from terminal zones by visual inspection of morphology in the 2D images (axons in terminal zones ramify and contain synaptic boutons, see Extended Data Fig. 6), it is difficult to confidently make this distinction algorithmically. We compared results of terminal labelling using Synaptophysin-EGFP-expressing AAV with the cytoplasmic EGFP AAV (Extended Data Fig. 6). Outside of major fibre tracts, there was high correspondence between synaptic EGFP and cytoplasmic EGFP signals in target regions. Nonetheless, it should be noted that the connectivity matrix contains passing fibre signals within grey matter, the nature of which should be manually examined in 2D section images.

This connectivity matrix (Fig. 3) has several striking features. First, connectivity strengths span a greater than 10^5-fold range across the brain (Extended Data Fig. 7), suggesting that quantitative descriptions
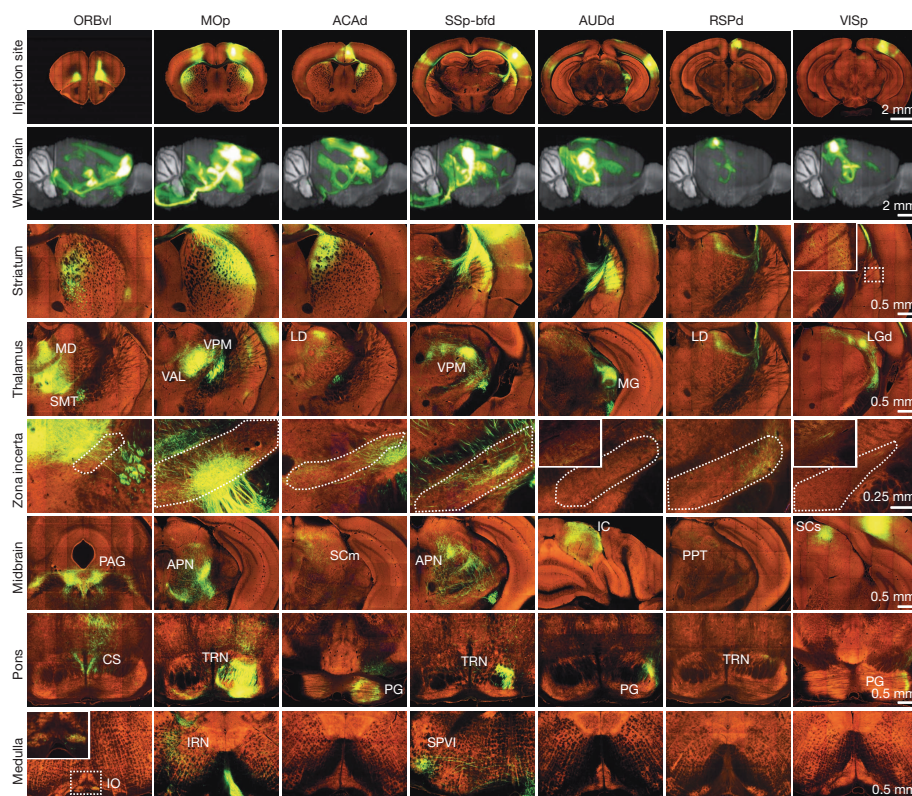
Figure 2 | Whole brain projection patterns from seven representative cortical regions. One coronal section at the centre of each injection site is shown in the top row (see Supplementary Table 1 for the full name of each region). In the second row, 3D thumbnails of signal density projected onto a sagittal view of the brain reveal differences in brain-wide projection patterns. The bottom 6 rows show examples of EGFP-labelled axons in representative subcortical regions.
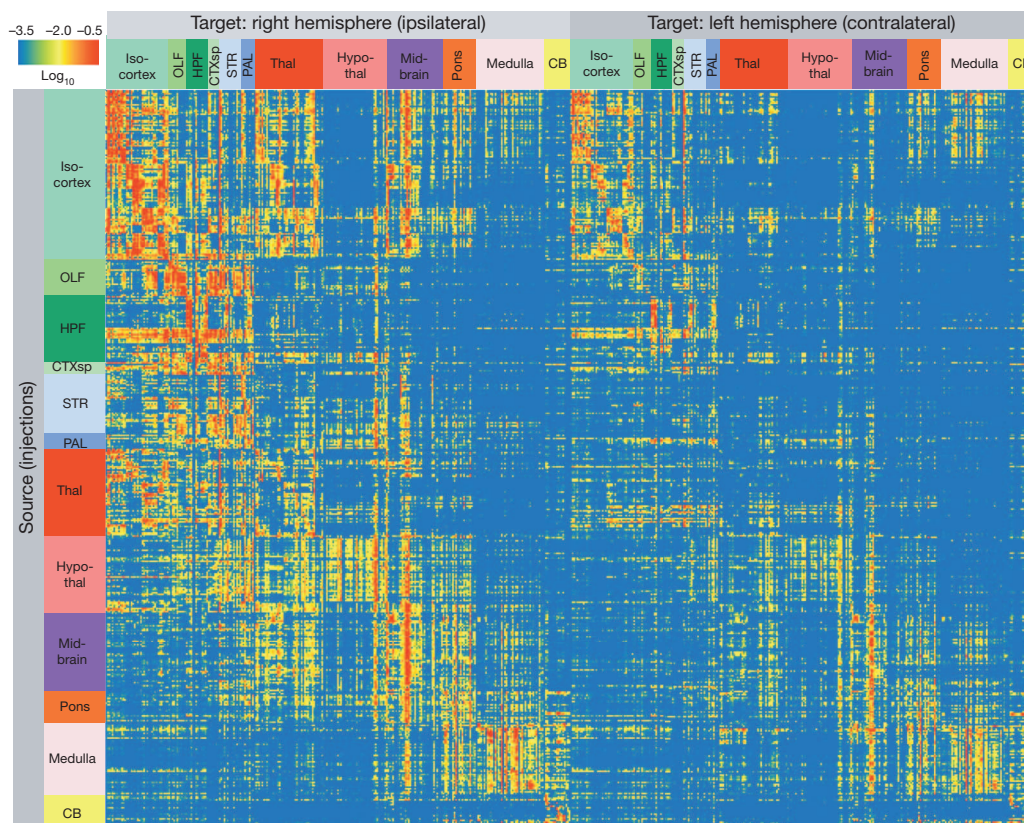
Figure 3 | Adult mouse brain connectivity matrix. Each row shows the quantitative projection signals from one of the 469 injected brains to each of the 295 non-overlapping target regions (in columns) in the right (ipsilateral) and left (contralateral) hemispheres. Both source and target regions are displayed in ontological order. The colour map indicates $\log_{10}$-transformed projection strength (raw values in Supplementary Table 2). All values less than $10^{-3.5}$ are shown as blue to minimize false positives due to minor tissue and segmentation artefacts and all values greater than $10^{-0.5}$ are shown as red to reduce the dominant effect of projection signals in certain disproportionately large regions (for example, striatum).

of connectivity must be considered for understanding neural network properties[24]. Second, there are prevalent bilateral projections to corresponding ipsilateral and contralateral target sites, with ipsilateral projections generally stronger than contralateral ones (total normalized projection volumes from all experiments are 4.3:1 between ipsilateral and contralateral hemispheres). Third, of all possible connections, strong connections are found in only a small fraction. Whereas 63% ipsilateral and 51% contralateral targets have projection strength values above the minimal true positive level of 10^{-4} (which has a potential false positive rate of 27%, Extended Data Fig. 7), only 21% ipsilateral and 9% contralateral targets have projection strength values above the intermediate level of 10^{-2}.

An inter-region connectivity model

Infected neurons in injection sites often span several brain areas. To better describe the mutual connection strengths between ontologically defined regions rather than injection sites, we constructed inter-region connectivity matrices via a computational model (see Supplementary Notes for a detailed description), using segmented projection volumes (Fig. 3) to define connection strengths. Two basic modelling assumptions were used. The first, regional homogeneity, assumes that projections between source X and target Y regions are homogeneously distributed, so that infection of a subarea of the region is representative of the entire region. This allows the value of $W_{X,Y}$, a regional connectivity measure, to be inferred from data that can at best only sample the source region. The second assumption, projection additivity, assumes that the projection density of multiple source regions sum linearly to produce projection density in a target region. This allows relative contributions of different sources to be determined for a target region, assuming at least partially independent injections.

The 469 experiments allowed us to compute the mutual connections among 213 regions. The best-fit model (Fig. 4a, see Supplementary Table 3 for the underlying values) results from a bounded optimization followed by a linear regression to determine connection coefficients, assigning statistical confidence (P values) to each connection in the matrix. Based on the bounded optimization, the number of non-zero entries provides an upper bound estimate for sparsity: 36% for the entire brain and 52% for cortico-cortical connections. Using confidence values for each non-zero connection, the lower bound on sparsity is 13% for the entire brain and 32% for cortico-cortical connections.

Connection strengths spanned $\sim 10^5$-fold range, and negatively correlated with the distance between connected regions (Supplementary Notes and Supplementary Table 4). Based on the Akaike information criterion (AIC), among hypothesized connection strength distributions (lognormal, normal, exponential, inverse Gaussian) the brain-wide data are best fit by a lognormal distribution (Fig. 4b, red lines). However, the log-transformed connection strengths failed to pass the Shapiro-Wilk test for normality (ipsilateral: $P = 0.039$; contralateral: $P = 0.023$), and among Gaussian mixture models, a two-component one provided the best fit (Fig. 4b, green lines). For cortico-cortical connections, both intra- and inter-hemispheric distributions are well fit by lognormals (ipsilateral: $P = 0.23$; contralateral: $P = 0.21$) individually (Fig. 4b), but they are different enough that when combined the distribution is no longer lognormal ($P = 0.0019$). This extends previously reported findings that cortico-cortical connections follow a lognormal distribution in the primate[24,25] and mouse cortex[26] to the entire mouse brain. These observations combined indicate that connections might be lognormally distributed within a region, yet vary systematically with statistics unique to the region.

Previous studies on connectivity considered global organizational principles from a graph-theory perspective[26–28]. We transformed our
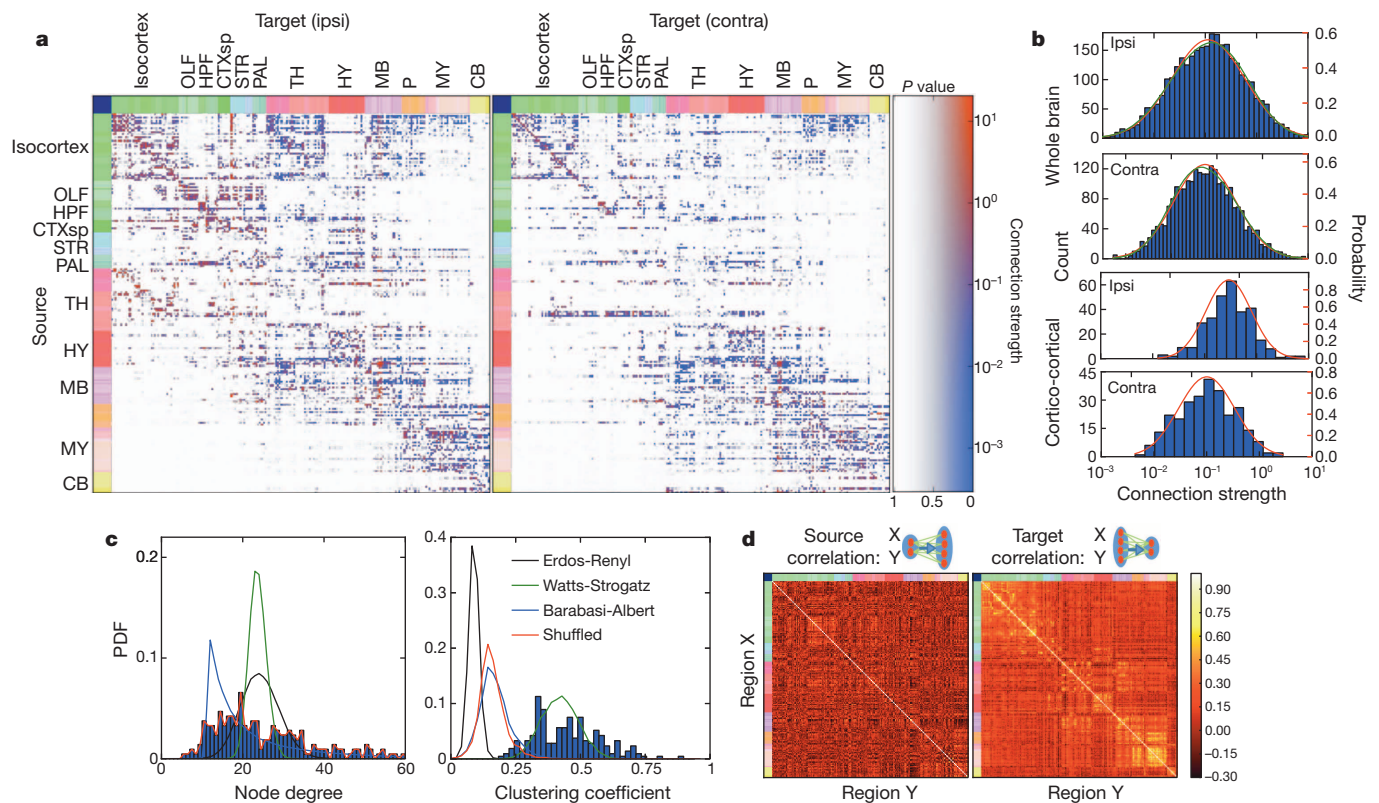
Figure 4 | A computational model of inter-regional connection strengths.
a, The inter-region connectivity matrix, with connection strengths represented in colours and statistical confidence depicted as an overlaid opacity. Note that in this matrix, the sources (rows) are regions, whereas for the matrix of Fig. 3, the sources are injection sites. **b**, Both whole-brain and cortico-cortical connections can be fit by one-component lognormal distributions (red lines). However, the log distributions of whole-brain connection strengths are best fit by a two-component Gaussian mixture model (green lines). **c**, Node degree and clustering coefficient distributions for a binarized version of the linear model network, compared against Erdos-Renyi, Watts-Strogatz and Barabasi-Albert networks with matched graph statistics. **d**, Comparison of the correlation coefficients of normalized connection density between areas, defined as the common source for projections to other regions (left) and as the common target of projections from other regions (right).

weighted, directed, connectivity matrix (Fig. 4a) to binary directed and binary undirected data sets. Network analyses (Fig. 4c, see also Supplementary Notes) reveal that the mouse brain has a higher mean clustering coefficient (which gives the ratio of existing over possible connections), 0.42, than expected by a random network[29,30] with identical sparseness, 0.12. Random graphs with matched node degree distribution show a similar drop in clustering coefficient to 0.16. A 'small-world' network model[31] approximates the clustering coefficient distribution after being fit to its mean; however, its node degree distribution poorly matches the data. Here, a better fit is achieved with a scale-free network[32]; however, neither model simultaneously fits both distributions.

Next, we analysed similarity in connection patterns between different regions. Similarity is characterized by two measures: correlations between outgoing projections originating in two areas and correlations between incoming projections ending in these two areas. Figure 4d depicts heat maps of correlation coefficients between the same regions of the linear model (Fig. 4a) depicted across the rows (that is, as a common source for other regions), and down the columns (that is, as a common target from other regions). The number of strong correlations is larger than expected by chance, suggesting a tendency of regions to organize into clusters to allow for strong indirect connectivity.

The cortico–striatal–thalamic network

Different cortical areas project to different domains of striatum and thalamus with some degree of topography[33–35]. We used 80 isocortical injection experiments to examine this. Spearman's rank correlation coefficient of segmented projection volumes of all voxels across the entire brain was computed between every pair of experiments, and hierarchical clustering led to 21 distinct groups, each containing 1 to 10 injections

(Extended Data Fig. 8a, b). Such grouping effectively divides the cortex into 21 predominantly non-overlapping spatial zones as shown in a flat-map cortex representation (Fig. 5a) defined by similar output projections. To effectively visualize different projection patterns in a common 3D space, voxel densities from 21 selected injections, one (centrally located) from each cluster, were overlaid to create 'dotograms' (Fig. 5b, c and Extended Data Fig. 8c), demonstrating that projections from different cortical regions divide up striatum and thalamus into distinct domains.

Average inter-group distances (Extended Data Fig. 9a–d) were used to quantify the degree to which inter-group spatial relationships within the cortex are preserved in target domains. Distance matrices for both ipsilateral and contralateral cortical targets were highly correlated with the distance matrix of injection sites, as were ipsilateral striatal and thalamic distance matrices. Weaker correlations were observed in contralateral striatum and thalamus. The computed distance matrices show that the spatial relationship between injection sites is recapitulated in the projections to striatum and thalamus, with some transformation of scale and rotation.

This highly synchronized topography can be determined via virtual tractography. Real tractography (following single axons) cannot be done because of the discrete 100-μm sampling between sections. Instead, from every voxel we computed a path back to the injection site by finding the shortest density-weighted distance through the voxels. The 3D tractography paths were plotted for both cortical hemispheres (Fig. 5d and Extended Data Fig. 9e), ipsilateral striatum and thalamus (Fig. 5e, f). The tractography shows that the paths themselves also retain the same spatial organization. In particular for the thalamus, anterior groups pass through fibre tracts in the striatum, narrowing through the globus pallidus,

before spreading throughout the thalamus (Extended Data Fig. 9f). Posterior groups (RSP, VIS) bypass the striatum but retain a strict topography following the medial/lateral axis (Fig. 5f).

Although the striatum is a cellularly homogeneous structure that can be subdivided into distinct domains selectively targeted by cortical and other inputs[36] (Fig. 5e), the thalamus is highly heterogeneous, composed of up to 50 discrete nuclei[37], receiving and relaying diverse sensory, motor, behavioural state and cognitive information in parallel pathways to and from the isocortex. We constructed a comprehensive wiring diagram between major, functionally distinct cortical regions and thalamic nuclei in the ipsilateral hemisphere (Fig. 6 and Extended Data Fig. 10), by combining the quantitative connectivity matrix (Fig. 3) with the linear model (Fig. 4a), manual proof-checking in the raw image sets, and cross-referencing published literature (83 publications, mostly from rat data, see Extended Data Fig. 10). This wiring diagram demonstrates specific point-to-point interconnections between corresponding clusters that divide the cortico-thalamic system into six functional pathways: visual, somatosensory, auditory, motor, limbic and prefrontal. We also observed cross-talk between these pathways, mediated by specific associational cortical areas and integrative thalamic nuclei.

The specific observations from our data are mostly consistent (with a few additions) with extensive previous studies in rats and the fewer number of studies in mice (Extended Data Fig. 10) as well as with studies in other mammalian species[37–39], providing a comprehensive and unifying view of mouse cortico-thalamic connections for the first time. Much work is still needed to obtain a full picture of connectivity in the cortico-thalamic system, including intra-cortical and intra-thalamic connections, their relationships with the interconnections between cortex

and thalamus, and the exquisite cortical laminar specificity of the originating and terminating zones of many of these connections[5,38,40].

Discussion

The standardized projection data set and the informatics framework built around it provide a brain-wide, detailed and quantitative connectivity map that is the most comprehensive, to date, in any vertebrate species. The high-throughput whole-brain mapping approach is remarkably consistent across animals, with an average correlation of 0.90 across 12 duplicate sets of mice (Extended Data Fig. 5). Informatics processing of the data set, such as co-registration and voxelization, helps with direct comparison between any image series, and systematic modelling and computational analyses of the entire network. Furthermore, the entire data set preserves the 3D spatial relationship of different domains, pathways and topography (Fig. 5). Thus, our connectivity atlas lays the groundwork for large-scale analyses of global neural networks, as well as networks within and between different neural systems.

As an initial analysis of this large-scale data set, we present an examination of both general principles of whole brain architecture and specific properties of cortical connections. We found that projections within the ipsilateral hemisphere and to the corresponding locations in the contralateral hemisphere are remarkably similar across the brain (Figs 3 and 4; Pearson's $r = 0.595$), with the contralateral connection strengths significantly weaker than ipsilateral ones. The mouse brain shows defining features of both small-world and scale-free networks, that is, it clusters and has hubs; but neither of these models in isolation can fully explain it. Interestingly, the connection strengths at both cortico-cortical and whole-brain levels show lognormal distributions, that is, long-tailed
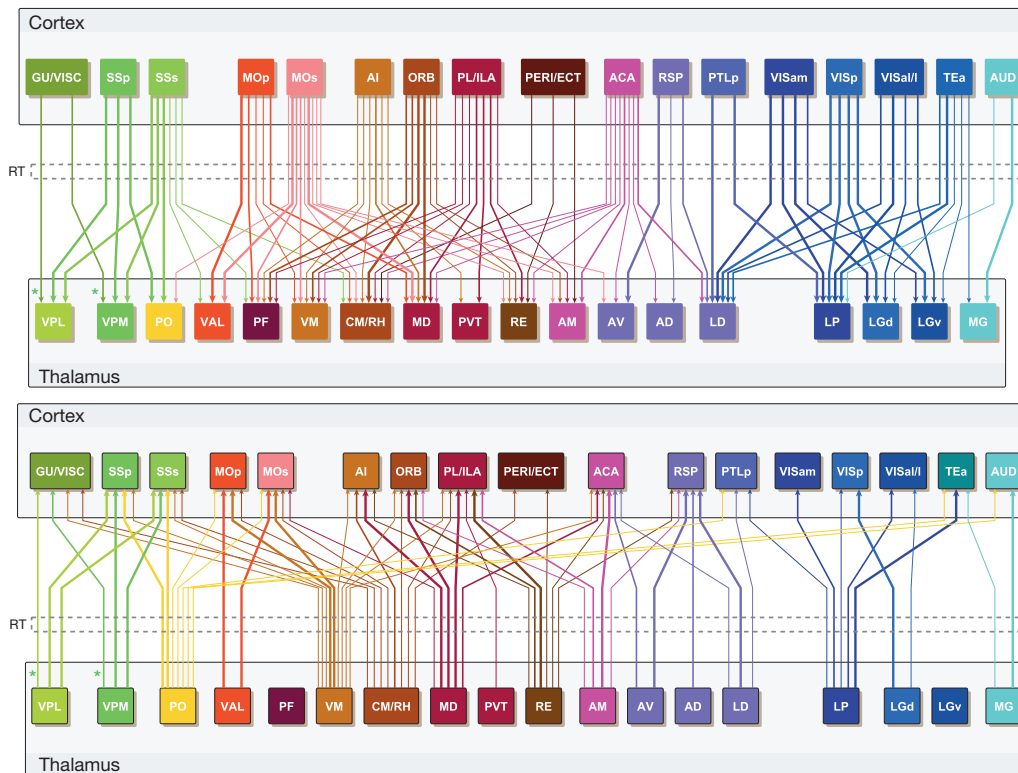
Figure 6 | A wiring diagram of connections between major cortical regions and thalamic nuclei. Upper and lower panels show projections from cortex to thalamus or from thalamus to cortex, respectively (ipsilateral projections only). Colour coding of different cortical regions and their corresponding thalamic nuclei is similar to the flat-map cortex in Fig. 5a. Thickness of the arrows indicates projection strength, which is shown in three levels as in Extended Data Fig. 10 and corresponds roughly to the red, orange and yellow colours in the raw connectivity matrix (Fig. 3). LGv and PF do not have significant projections to cortex. The reticular nucleus of the thalamus (RT) (the dashed box) is placed in between cortex and thalamus to illustrate its special role as a relay nucleus which all cortico-thalamic and thalamo-cortical projections pass through and make collateral projections into. The asterisks indicate that cortico-thalamic and thalamo-cortical projections in the gustatory/visceral pathway are between GU/VISC cortical areas and VPMpc/VPLpc nuclei (instead of VPM/VPL). See Supplementary Table 1 for the full name of each region.

distributions with small numbers of strong connections and large numbers of weak connections. In connections among isocortex, striatum and thalamus, clustering analysis and virtual tractography recapitulate anatomical parcellation and topography of functional domains and projection pathways (Fig. 5). The extensive reciprocal connections between isocortex and thalamus (Fig. 6) further illustrate general principles of network segregation and integration.

Our Connectivity Atlas represents a first systematic step towards the full understanding of the complex connectivity in the mammalian brain. Through the process, limitations of the current approach and opportunities for future improvement can be identified. On the technical side, any potential new connections identified in the Phase I data set (which does not yet have extensive redundancy in regional coverage) will need to be confirmed with more data. Also, we cannot exclude the possibility that the AAV tracer we chose to use (with the specific promoter and serotype) may not be completely unbiased in labelling all neuronal types. The connectivity matrix has been shown to contain false positive signals (Extended Data Fig. 7), mainly due to tissue and imaging artefacts and injection tract contaminations. The connectivity matrix based on cytoplasmic EGFP labelling does not distinguish passing fibres from terminal zones, and examination of raw images is needed to help with such distinction, using features such as ramification of axon fibres, and boutons or enlargements in axons. The Atlas could also be enhanced in the future with more systematic mapping using synaptic-terminal-specific viral tracers as shown (Extended Data Fig. 6). Regarding signal quantification, we chose to use projection volume (sum of segmented pixel counts) over projection fluorescence intensity (sum of segmented pixel intensity), because we found the former more reliable and less variable across different brains (even after normalization). However, the use of projection volume will probably underestimate the strength of dense projections. Thus, the true range of projection strengths may go beyond the 10^5-fold reported here. Finally, we observed that the alignment between the average template brain and our existing Reference Atlas model (which was drawn upon Nissl sections) is not perfect, which

leads to a degree of registration imprecision that could affect the accuracy of the quantitative connectivity matrices. Our work shows the need to generate a new reference model based on a realistic 3D brain, such as the average template brain presented here. Our data set can also help this by adding connectivity information to improve anatomical delineations previously defined solely by cyto- and chemoarchitecture.

Beyond the above technical issues, identities of the postsynaptic neurons at the receiving end of the mapped connections are not labelled and therefore unknown. Microscale, synaptic-level details are missing, and electrical connections through gap junctions are not revealed. Moreover, our mesoscale connectome provides a static, structural connectivity map, which is necessary but insufficient for understanding function. Moving from here to functional connectivity and circuit dynamics in a living brain will require fundamentally different approaches[41]. One important aspect concerns the types of synapses present in each connectional path, as determined by their neurotransmitter contents and their physiological properties. Anatomical connection strength (for example, numbers of axon fibres and boutons) needs to be combined with physiological connection properties (for example, excitatory vs inhibitory types of synapses, fast vs slow neurotransmission, and the specific strength and plasticity of each synapse) to yield a true functional connection strength.

With the goal of bridging structural connectivity and circuit function[41], we have taken a genetic approach, using AAV viral tracers that express EGFP in either a pan-neuronal or cell-type-specific manner. The same neural networks mapped here can be further investigated by similar viral vectors expressing tools for activity monitoring (for example, genetically encoded calcium indicators) and activity manipulation (for example, channelrhodopsins)[16]. Furthermore, our ongoing efforts of Cre-driver dependent tracing will allow more specific connectivity mapping from discrete areas and specific functional cell types. Such cell-type-specific connectivity mapping is perhaps the greatest advantage of the genetic tracing approach, allowing dissection of differential projection patterns from different neuronal types that are often intermingled in the same region. The genetic tracing approach can be further extended

into identification of inter-connected pre- and postsynaptic cell types and individual cells, using approaches such as retrograde or antero-grade trans-synaptic tracing[42–44]. Our approach can also be applied to animal models of human brain diseases and the connectivity data generated here can be instructive to human connectome studies, which will help to further our understanding of human brain connectivity and its involvement in brain disorders.

METHODS SUMMARY

C57BL/6J male mice at age P56 were injected with EGFP-expressing AAV using iontophoresis by the stereotaxic method. The brains were scanned using the STP tomography systems. Images were subject to data quality control and all the injection sites were manually annotated. All the image sets were co-registered into the 3D reference space. EGFP-positive signals were segmented from background, and binned at voxel levels for quantitative analyses. The raw connectivity data are served with various navigation tools on the web through the Allen Institute's data portal. See the full Methods section for detailed descriptions.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Swanson, L. W. *Brain Architecture: Understanding the Basic Plan* 331 (Oxford Univ. Press, 2012).
2. Seung, S. *Connectome: How The Brain's Wiring Makes Us Who We Are* 359 (Houghton Mifflin Harcourt, 2012).
3. Sporns, O. *Networks of the Brain* 412 (The MIT Press, 2011).
4. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans. Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
5. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
6. Bota, M., Dong, H. W. & Swanson, L. W. Combining collation and annotation efforts toward completion of the rat and mouse connectomes in BAMS. *Front. Neuroinform.* **6**, 2 (2012).
7. Chiang, A. S. *et al.* Three-dimensional reconstruction of brain-wide wiring networks in *Drosophila* at single-cell resolution. *Curr. Biol.* **21**, 1–11 (2011).
8. Jenett, A. *et al.* A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep.* **2**, 991–1001 (2012).
9. Sporns, O., Tononi, G. & Kotter, R. The human connectome: A structural description of the human brain. *PLoS Comput. Biol.* **1**, e42 (2005).
10. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: an overview. *NeuroImage* **80**, 62–79 (2013).
11. Bock, D. D. *et al.* Network anatomy and *in vivo* physiology of visual cortical neurons. *Nature* **471**, 177–182 (2011).
12. Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471**, 183–188 (2011).
13. Chklovskii, D. B., Vitaladevuni, S. & Scheffer, L. K. Semi-automated reconstruction of neural circuits using electron microscopy. *Curr. Opin. Neurobiol.* **20**, 667–675 (2010).
14. Helmstaedter, M. *et al.* Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500**, 168–174 (2013).
15. Takemura, S. Y. *et al.* A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature* **500**, 175–181 (2013).
16. Huang, Z. J. & Zeng, H. Genetic approaches to neural circuits in the mouse. *Annu. Rev. Neurosci.* **36**, 183–215 (2013).
17. Masland, R. H. The neuronal organization of the retina. *Neuron* **76**, 266–280 (2012).
18. Bohland, J. W. *et al.* A proposal for a coordinated effort for the determination of brainwide neuroanatomical connectivity in model organisms at a mesoscopic scale. *PLoS Comput. Biol.* **5**, e1000334 (2009).
19. Chamberlin, N. L., Du, B., de Lacalle, S. & Saper, C. B. Recombinant adeno-associated virus vector: use for transgene expression and anterograde tract tracing in the CNS. *Brain Res.* **793**, 169–175 (1998).
20. Harris, J. A., Oh, S. W. & Zeng, H. Adeno-associated viral vectors for anterograde axonal tracing with fluorescent proteins in nontransgenic and Cre driver mice. *Curr. Protoc. Neurosci.* **59**, 1.20.1–1.20. 18 http://dx.doi.org/10.1002/0471142301.ns0120s59 (2012).
21. Wang, Q. *et al.* Systematic comparison of adeno-associated virus and biotinylated dextran amine reveals equivalent sensitivity between tracers and novel projection targets in the mouse brain. *J. Comp. Neurol.* http://dx.doi.org/10.1002/cne.23567 (Feb 22, 2014).
22. Ragan, T. *et al.* Serial two-photon tomography for automated *ex vivo* mouse brain imaging. *Nature Methods* **9**, 255–258 (2012).
23. Dong, H. W. *The Allen Reference Atlas: a Digital Color Brain Atlas of the C57BL/6J Male Mouse* (John Wiley & Sons, 2008).
24. Markov, N. T. *et al.* A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb. Cortex* **24**, 17–36 (2014).
25. Ercsey-Ravasz, M. *et al.* A predictive network model of cerebral cortical connectivity based on a distance rule. *Neuron* **80**, 184–197 (2013).
26. Wang, Q., Sporns, O. & Burkhalter, A. Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *J. Neurosci.* **32**, 4386–4399 (2012).
27. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Rev. Neurosci.* **10**, 186–198 (2009).
28. Schmitt, O. *et al.* The intrinsic connectome of the rat amygdala. *Front. Neural Circuits* **6**, 81 (2012).
29. Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–61 ((1960).
30. Bollobás, B. *Random Graphs* (Cambridge Univ. Press, 2001).
31. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
32. Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
33. Groenewegen, H. J. & Berendse, H. W. The specificity of the 'nonspecific' midline and intralaminar thalamic nuclei. *Trends Neurosci.* **17**, 52–57 (1994).
34. Joel, D. & Weiner, I. The organization of the basal ganglia-thalamocortical circuits: open interconnected rather than closed segregated. *Neuroscience* **63**, 363–379 (1994).
35. Mailly, P., Aliane, V., Groenewegen, H. J., Haber, S. N. & Deniau, J. M. The rat prefrontostriatal system analyzed in 3D: evidence for multiple interacting functional units. *J. Neurosci.* **33**, 5718–5727 (2013).
36. Gerfen, C. R. in *The Rat Nervous System* (ed. Paxinos, G.) 455–508 (Elsevier Academic Press, 2004).
37. Jones, E. G. *The Thalamus* (Cambridge Univ. Press, 2007).
38. Jones, E. G. Viewpoint: the core and matrix of thalamic organization. *Neuroscience* **85**, 331–345 (1998).
39. Scannell, J. W., Burns, G. A., Hilgetag, C. C., O'Neil, M. A. & Young, M. P. The connectional organization of the cortico-thalamic system of the cat. *Cereb. Cortex* **9**, 277–299 (1999).
40. Barbas, H. & Rempel-Clower, N. Cortical structure predicts the pattern of corticocortical connections. *Cereb. Cortex* **7**, 635–646 (1997).
41. Bargmann, C. I. & Marder, E. From the connectome to brain function. *Nature Methods* **10**, 483–490 (2013).
42. Lo, L. & Anderson, D. J. A Cre-dependent, anterograde transsynaptic viral tracer for mapping output pathways of genetically marked neurons. *Neuron* **72**, 938–950 (2011).
43. Wall, N. R., Wickersham, I. R., Cetin, A., De La Parra, M. & Callaway, E. M. Monosynaptic circuit tracing *in vivo* through Cre-dependent targeting and complementation of modified rabies virus. *Proc. Natl Acad. Sci. USA* **107**, 21848–21853 (2010).
44. Wickersham, I. R. *et al.* Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53**, 639–647 (2007).

Author Information The Allen Mouse Brain Connectivity Atlas is accessible at (http://connectivity.brain-map.org). All AAV viral tracers are available at Penn Vector Core, and AAV viral vector DNA constructs have been deposited at the plasmid repository Addgene. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.Z. (hongkuiz@alleninstitute.org).

ARTICLE

MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool

Helge Gad[1]*, Tobias Koolmeister[1]*, Ann-Sofie Jemth[1]*, Saeed Eshtad[1]*, Sylvain A. Jacques[1]*, Cecilia E. Ström[1]*, Linda M. Svensson[2], Niklas Schultz[1], Thomas Lundbäck[1,3], Berglind Osk Einarsdottir[4], Aljona Saleh[5], Camilla Göktürk[1], Pawel Baranczewski[1,6], Richard Svensson[3,6], Ronnie P.-A. Berntsson[2], Robert Gustafsson[2], Kia Strömberg[1], Kumar Sanjiv[1], Marie-Caroline Jacques-Cordonnier[1], Matthieu Desroses[1], Anna-Lena Gustavsson[1,3], Roger Olofsson[4], Fredrik Johansson[7], Evert J. Homan[1], Olga Loseva[1], Lars Bräutigam[1], Lars Johansson[1,3], Andreas Höglund[1], Anna Hagenkort[1], Therese Pham[1], Mikael Altun[1], Fabienne Z. Gaugaz[1,6], Svante Vikingsson[1,8], Bastiaan Evers[1]†, Martin Henriksson[1], Karl S. A. Vallin[1], Olov A. Wallner[1], Lars G. J. Hammarström[1,3], Elisee Wiita[1], Ingrid Almlöf[1], Christina Kalderén[1], Hanna Axelsson[1,3], Tatjana Djureinovic[7]†, Jordi Carreras Puigvert[1], Maria Häggblad[9], Fredrik Jeppsson[1], Ulf Martens[9], Cecilia Lundin[1], Bo Lundgren[9], Ingrid Granelli[5], Annika Jenmalm Jensen[1,3], Per Artursson[3,6], Jonas A. Nilsson[4], Pål Stenmark[2], Martin Scobie[1], Ulrika Warpman Berglund[1] & Thomas Helleday[1]

Cancers have dysfunctional redox regulation resulting in reactive oxygen species production, damaging both DNA and free dNTPs. The MTH1 protein sanitizes oxidized dNTP pools to prevent incorporation of damaged bases during DNA replication. Although MTH1 is non-essential in normal cells, we show that cancer cells require MTH1 activity to avoid incorporation of oxidized dNTPs, resulting in DNA damage and cell death. We validate MTH1 as an anticancer target *in vivo* and describe small molecules TH287 and TH588 as first-in-class nudix hydrolase family inhibitors that potently and selectively engage and inhibit the MTH1 protein in cells. Protein co-crystal structures demonstrate that the inhibitors bind in the active site of MTH1. The inhibitors cause incorporation of oxidized dNTPs in cancer cells, leading to DNA damage, cytotoxicity and therapeutic responses in patient-derived mouse xenografts. This study exemplifies the non-oncogene addiction concept for anticancer treatment and validates MTH1 as being cancer phenotypic lethal.

Anticancer treatments are dominated by targeting the genetic defects found in cancers, such as oncogenes (for example, imatinib for *BCR-ABL* mutated cancer[1]) or non-oncogenic genetic defects (for example, PARP inhibitors for *BRCA1* and *BRCA2* mutated cancer[2,3]). Although a few cases show impressive results, these treatments have, in general, not replaced radiotherapy or chemotherapy. Targeting genetic defects in a personalized strategy is limited by the high degree of intra-tumour heterogeneity, adaptation of genetic networks and high somatic mutation rates in cancer[4]. Here, we reasoned that a de-personalized anticancer strategy, targeting the cancer phenotype using a synthetic-lethal approach, may tackle the problem of intra-tumour heterogeneity and be more widely applicable to a range of tumours.

Altered redox regulation is a general phenotype of many cancers and has a crucial role in cancer aetiology, progression and metastasis[5]. Dysfunctional redox regulation and an increased reactive oxygen species (ROS) tension can cause oxidative damage to DNA directly, or to free bases in the cellular and mitochondrial deoxynucleoside triphosphate (dNTP) pool[6]. Damage on double-stranded DNA has been the main therapeutic focus in spite of the fact that the free dNTP precursor pool is 190–13,000 times more susceptible to damage[7]. The MTH1 protein sanitizes the oxidized dNTP pool, for example, by converting 8-oxodGTP or 2-OH-dATP into 8-oxodGMP or 2-OH-dAMP, thus avoiding incorporation of these oxidized nucleotides into the DNA, which otherwise

can result in mispairing, mutations and cell death[8–10]. Interestingly, overexpression of MTH1 suppresses the mutator phenotype in mismatch-repair-defective colorectal cancer cells[11], indicating that the oxidized dNTP pool is a major source of damage causing mutations in cancer. Furthermore, overexpression of MTH1 reverses Ras-induced senescence by suppressing the overall level of DNA damage[12], underscoring the importance of oxidized dNTPs in cancer. As altered redox status can be exploited for cancer treatments[5], we reasoned that the normally non-essential MTH1 protein[13] may be important for sanitizing cancer-associated damage in the dNTP pool and required for survival in cancer cells. Here, we wanted to evaluate whether MTH1 could be a general target for cancer treatment owing to non-oncogene addiction; that is, cancer phenotypic lethal.

MTH1 is required for cancer survival

To investigate MTH1 as a target for cancer treatment, the protein was depleted using three different short interfering RNAs (siRNAs) in several cancer cell lines, resulting in DNA damage and reduced clonogenic survival and viability; the variable degree of responses are probably due to the transient and incomplete depletion of the protein (Extended Data Fig. 1). Furthermore, rescue expression of the RNA interference (RNAi)-resistant wild-type protein, but not MTH1(E56A) catalytically dead protein, fully reversed the reduced survival in U2OS cells (Fig. 1a and Extended

Data Fig. 2a–c). Notably, no attenuation of survival was observed in primary VH10 cells. These cells display a lower level of carbonylated proteins, indicative of lower levels of ROS, compared to U2OS cancer cells (Extended Data Fig. 2j), in agreement with MTH1 being non-essential in normal cells and knockout mice having mild phenotypes[13]. MTH1 depletion in cancer cells results in the accumulation of 8-oxodG in DNA and nuclear 53BP1 foci, occurring at DNA double-strand breaks (DSBs)[14]. These effects are fully reversed by expression of an RNAi-resistant wild-type MTH1 but not the catalytically dead protein (Fig. 1b, c). Although 8-oxodG incorporation couldn't be confirmed by liquid chromatography-mass spectrometry (LC-MS/MS) owing to the widely reported
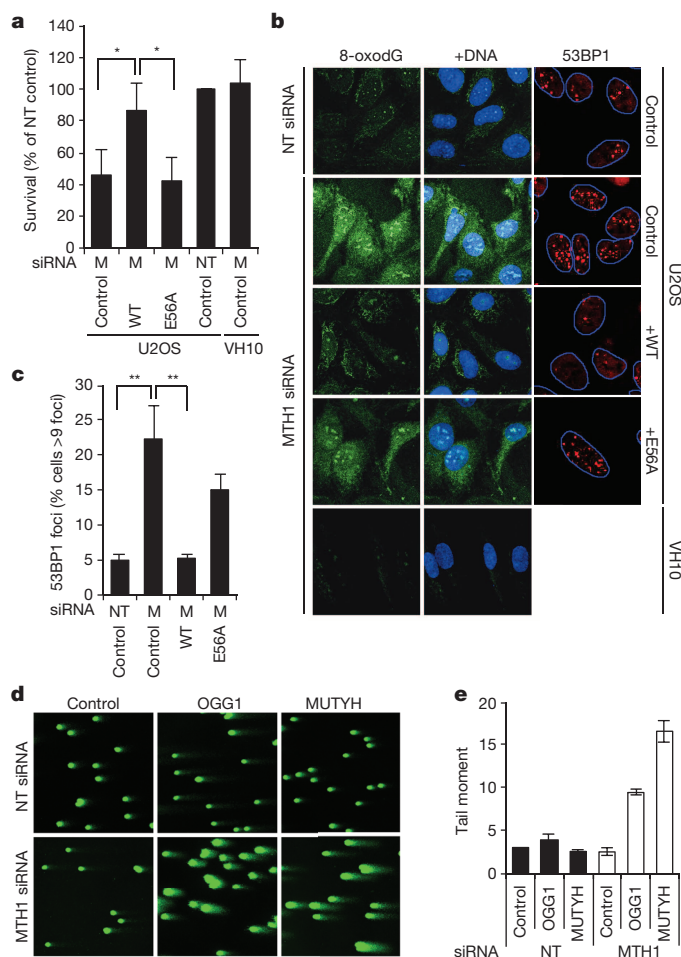
Figure 1 | MTH1 prevents 8-oxodGTP incorporation and is required for cancer cell survival. a, Clonogenic survival of U2OS and VH10 cells transiently transfected with MTH1 siRNA (M) or non-targeting siRNA (NT). Overexpression of wild-type (WT) MTH1 but not the catalytically dead mutant (E56A) reverses the decrease in survival after MTH1 knockdown. Data shown as average ± s.d. from three independent experiments. *$P < 0.05$, one-way analysis of variance (ANOVA). **b**, 8-oxodGTP incorporation in DNA (Avidin-AlexaFluor488 reactive substance) and induction of 53BP1 foci in U2OS and VH10 cells transiently transfected with non-targeting siRNA or *MTH1* siRNA for 6 days. Overexpression of wild-type MTH1 but not the catalytically dead mutant (E56A) rescues both 8-oxodG and 53BP1 foci formation. **c**, Quantification of 53BP1 foci formation in U2OS cells with overexpression of wild-type and catalytically dead mutant (E56A) after siRNA transfection for 6 days. Data shown as average ± s.d. from three independent experiments (≥200 cells per sample size). **$P < 0.01$, one-way ANOVA. **d**, *MTH1* siRNA depletion increases 8-oxodG and 2-OH-dA levels in DNA. Alkaline comets of U2OS cells transfected with non-targeting siRNA or *MTH1* siRNA no. 1, incubated with OGG1, MUTYH or buffer alone (control) for 45 min at 37 °C. **e**, Quantification of comet tail moment. Values represent average ± s.e.m. from two independent experiments (>100 comets per experiment).

generation of oxidized bases during sample preparation[15], we were able to detect 8-oxodG and 2-OH-dA in DNA by the modified Comet assay (Fig. 1d, e), using OGG1 (recognizing 8-oxodG) and MUTYH (recognizing 8-oxodG and 2-OH-dA). These data indicate that MTH1 prevents incorporation of both 8-oxodG and 2-OH-dA (ref. 16). To determine whether 8-oxodG and/or 2-OH-dA contribute to cytotoxicity after MTH1 loss, we expressed RNAi-resistant MTH1 mutants D119A and W117Y, defective in 8-oxodGTPase and 2-OH-dATPase activity, respectively[16], neither of which rescued viability (Extended Data Fig. 2g–i). These data indicate that both 8-oxodG and 2-OH-dA contribute to the cytotoxicity.

DNA damage induced by MTH1 depletion activates RAD51- and DNA-PKcs-mediated DSB repair (Fig. 2), as well as ATM-dependent phosphorylation of p53 (S15) and upregulated expression of the p53 target gene *p21* (Fig. 2a, g and Extended Data Fig. 2k, l). p53 is known to mediate apoptosis after DNA damage, and short hairpin RNA (shRNA)-mediated depletion of p53 in U2OS cells prevented p21 induction and apoptosis after siRNA-mediated depletion of MTH1 (Fig. 2g–i). However, DNA damage is also induced in p53-depleted cells and the p53 status does not influence trimethylation of lysine 9 in histone H3 (H3K9me3), which is a marker for senescence[17] (Fig. 2h).

We next established *TP53* mutant SW480 colorectal cancer cells stably carrying a doxycycline-inducible shRNA vector targeting *MTH1*, which when induced, reduced MTH1 expression and clonogenic survival (Fig. 2j, k). In mice, tumours with the *MTH1* shRNA construct all responded when doxycycline was added to the drinking water, resulting in improved survival (Fig. 2l, m and Extended Data Fig. 2m, n). Doxycycline treatment of mice carrying tumours expressing non-targeting shRNA had no effect on tumour growth.

Development of MTH1 inhibitors

Our data demonstrate that MTH1 catalytic activity is required for cancer cell survival, as only the RNAi-resistant wild-type MTH1, and not the catalytically dead MTH1(E56A), restores survival. This suggests that similar effects could be obtained by MTH1 inhibition. We purified the MTH1 protein, screened compound libraries and identified small molecule hits containing a 2-aminopyrimidine motif (scaffold 1) that efficiently inhibited MTH1 catalytic activity. Introduction of an aminomethyl substituent TH086 (**1**) provided a marked increase in potency (Extended Data Fig. 3b–e). After hit expansion, TH287 (**2**) was identified as a potent MTH1 inhibitor (half-maximum inhibitory concentration (IC_{50}) = 0.8 ± 0.1 nM; Extended Data Fig. 3).

TH287 was rapidly metabolized in human and mouse liver microsomes both *in vitro* and *in vivo* via *N*-dealkylation of the aminomethyl substituent (Extended Data Table 2 and Extended Data Fig. 4a). Replacement of the methyl group by a cyclopropyl substituent in TH588 (**3**) improved metabolic stability both *in vitro* and *in vivo*, while maintaining MTH1 potency (IC_{50} = 5.0 ± 0.2 nM) (Extended Data Figs 3 and 4b, c and Extended Data Table 2). In contrast, introduction of an oxetanyl ring in TH650 (**4**) also improved metabolic stability but markedly reduced MTH1 potency (IC_{50} = 2.1 ± 0.1 μM). Using a target engagement assay[18] we observed that MTH1 inhibitors TH287 and TH588, but not TH650, bind the MTH1 protein in cells.

MTH1 inhibitor co–crystals

To gain further insight into the mechanism of how TH287 and TH588 inhibit MTH1, we determined the structures of MTH1 in complex with TH287 and TH588 to a resolution of 1.6 Å (Fig. 3, Extended Data Table 1 and Extended Data Fig. 3i, j). The aminopyrimidine moiety in TH287 and TH588 binds in the active site with key hydrogen bonds to Asn 33, Asp 119 and Asp 120 (Fig. 3b–d). As a result of the cyclopropyl substitution in TH588 the dichlorophenyl ring has rotated approximately 180° between TH287 and TH588. The aminomethyl substituent in TH287 and the corresponding amino group in TH588 appear to be a feature that is important for affinity and specificity, and we speculate that the
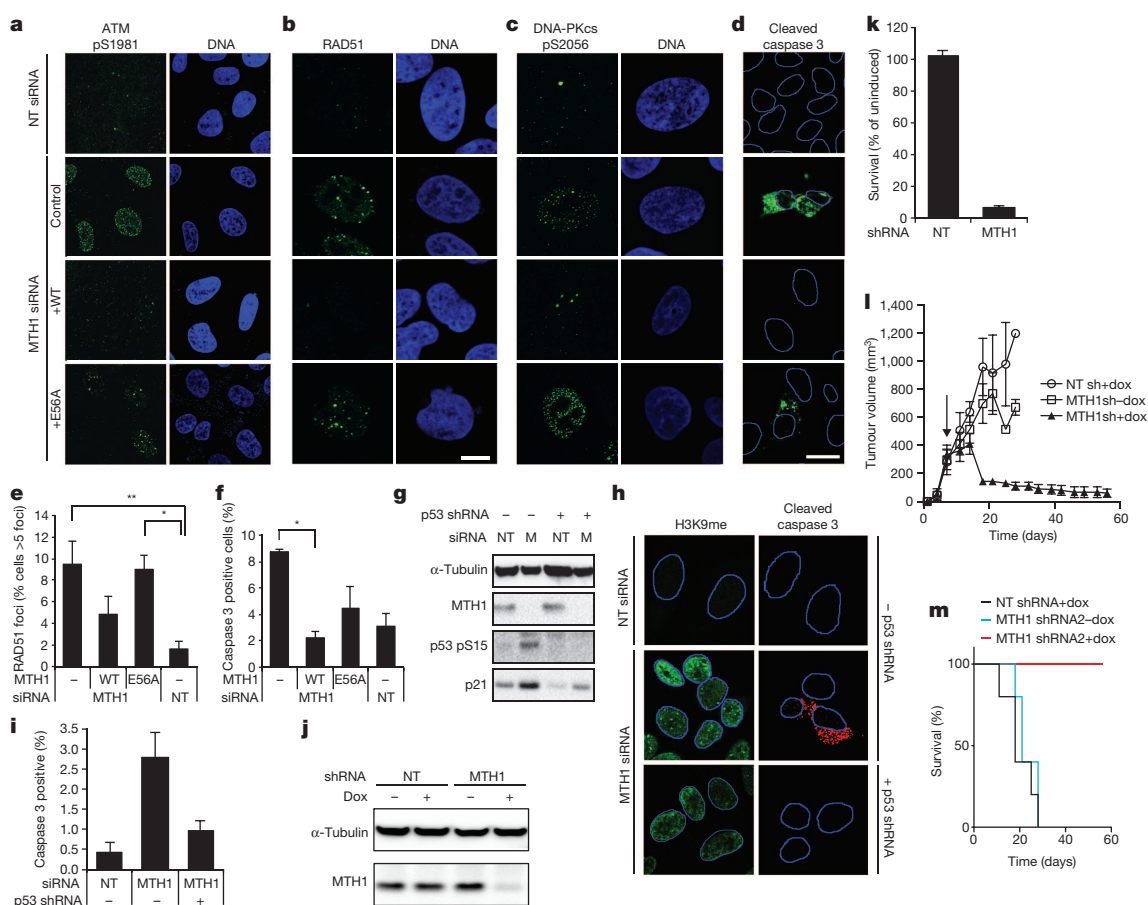
Figure 2 | MTH1 activity is required to prevent DNA damage and promote tumour growth in xenograft mice. a–d, After *MTH1* siRNA treatment, ATM autophosphorylation at serine 1981 (**a**), RAD51 foci (**b**), phosphorylated DNA-PKcs pS2056 foci (**c**) and caspase 3 cleavage (apoptosis) (**d**) were visualized. Expression of siRNA-resistant MTH1 wild-type prevents ATM activation, RAD51 foci, DNA-PKcs pS2056 foci and apoptosis but not expression of MTH1 catalytically dead mutant (E56A). Nuclei are indicated with thin blue lines. Scale bar, 10 μm. **e,** Quantification of RAD51 foci. Data shown as average ± s.e.m. from three independent experiments (≥200 cells per sample). *$P < 0.05$, **$P < 0.01$, one-way ANOVA. **f,** Quantification of cleaved-caspase-3-positive cells from panel **d**. Data shown as average ± s.e.m. from two independent experiments ($P < 0.05$, one-way ANOVA). **g, h,** Western blot (**g**) and immunofluorescence (**h**) of U2OS cells depleted of p53 by doxycycline-inducible shRNA expression (*p53* shRNA) and transfected with non-targeting siRNA (NT) or *MTH1* siRNA (M). **h,** Images of cells immunostained for

apoptosis (cleaved caspase 3) and senescence (H3K9me3). Nuclei are indicated with thin blue lines. **i,** Quantification of cleaved caspase 3 cells from panel **h**. Data shown as average ± s.d. from two independent experiments. **j,** Western blot of protein lysates isolated from SW480 cells with either non-targeting shRNA or *MTH1* shRNA after 72-h treatments with 2 μg ml^{-1} doxycycline (dox) in the medium. **k,** Clonogenic survival of SW480 cells of the cells in panel **j**. Cells were treated with doxycycline for 72 h and re-seeded for colony outgrowth. Values are indicated as survival compared to non-doxycycline-treated control cells and represent average ± s.d. from three independent experiments. **l, m,** SW480 cells expressing doxycycline-inducible *MTH1* shRNA or non-targeting shRNA were injected into SCID mice (subcutaneous) ($n = 5$ per group). After 7 days, doxycycline was added (arrow) to the drinking water. **l, m,** Tumour volume decreased (**l**) and survival increased in xenograft mice (**m**) with *MTH1* shRNA depletion.

hydrogen bond between NH and Asp 119 could mimic the 6-enol state of 8-oxodGTP suggested previously to be important for specificity[16,19].

MTH1 inhibitors cause cancer cell death

We observed that the MTH1 inhibitors TH287 and TH588 selectively and effectively kill U2OS and other cancer cell lines, but were considerably less toxic to several primary or immortalized cells (Fig. 4c and Extended Data Fig. 5j–m). This is in agreement with MTH1 being required for cancer cell survival while being non-essential in untransformed cells. The structurally similar compound TH650, which does not engage or inhibit the MTH1 protein in cells, is not toxic in any cells tested, supporting inhibition of MTH1 as a mechanism of induced cytotoxicity for TH287 and TH588 (Extended Data Fig. 5n). Notably, although the MTH1 inhibitors are low-nanomolar enzymatic inhibitors they are only cytotoxic at low micromolar concentrations. This is not an uncommon phenomenon and may be explained by pharmacokinetic and/or pharmacological properties such as high protein binding or low permeability/efflux in cancer cells (Extended Data Table 2).

Compounds that exert their toxicity through MTH1 inhibition would be expected to increase 8-oxodG incorporation into DNA. Indeed, we observed an increase of 8-oxodG in DNA after TH287 and TH588 treatment in U2OS cells, but not in VH10 cells or after treatment with TH650 (Fig. 4d–g and Extended Data Fig. 5a, c), in line with our proposed mechanism of action. Also, MTH1 inhibitors TH287 and TH588 both induce DNA damage (53BP1, RPA and phospho-DNA-PKcs foci) and trigger an ATM-p53-mediated death response and DNA repair in U2OS cells, but not in VH10 cells (Fig. 4h and Extended Data Fig. 6). However, p53 status does not influence survival on treatment with either TH287 or TH588 (Extended Data Fig. 5f–i), which is explained by DNA damage being introduced independently of p53 status. Overall, the MTH1 inhibitors induced a similar DNA damage response and cell death mechanism as observed after *MTH1* siRNA depletion. Although the OGG1 and MUTYH proteins are important in repair of 8-oxodG and 2-OH-dA lesions in DNA[20], differential expression levels or over-expression of either OGG1 or MUTYH did not alter the cytotoxic effect of MTH1 inhibitors (Extended Data Figs 7 and 8f), which is potentially
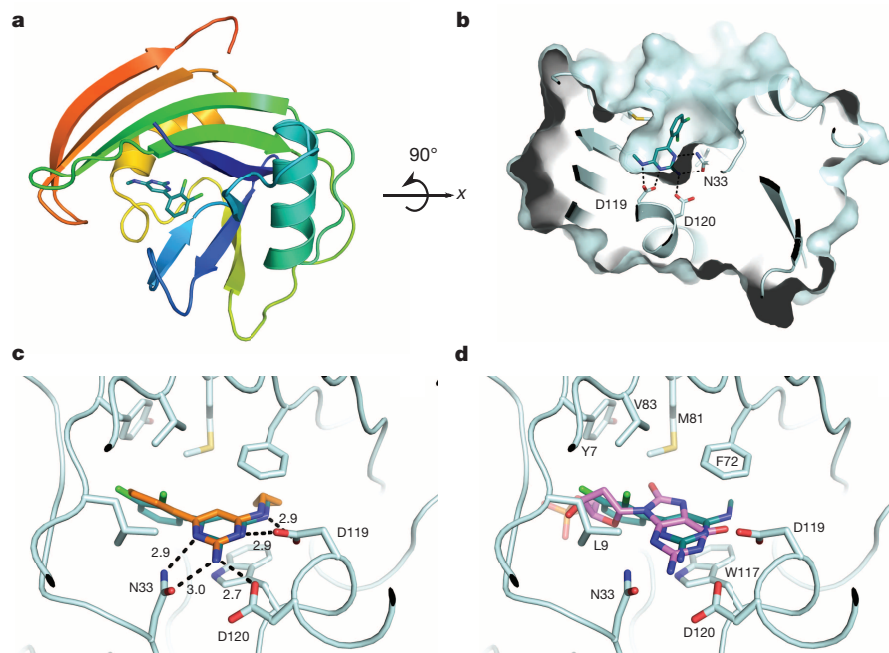
Figure 3 | Structural details of TH287 binding to MTH1. a, Overall structure of MTH1 shown in cartoon representation and coloured blue to red from amino to carboxy terminus. The bound TH287 inhibitor is shown in sticks representation. **b**, Sliced surface representation, highlighting the deep pocket of MTH1 (cyan) where TH287 (dark teal) is bound. The three residues that form hydrogen bonds to TH287 are marked in the figure. **c**, Close-up view of the binding site, comparing the TH287 (dark teal) ligand with the TH588 (orange) ligand, with the hydrogen bonding distance marked out in Å. **d**, Comparison of TH287 (dark teal) and 8-oxodGMP (violet) binding to MTH1, with important residues in the binding site marked.

explained by an overload of oxidized bases in DNA, overwhelming the base excision repair system, or that 8-oxodGTP or 2-OH-dATP may potentially signal for cell death[21]. Also, the expression levels of OGG1, MUTYH or MTH2 were unaltered after MTH1 depletion (Extended Data Figs 8b–d).

To determine at which point in the cancer transformation process cells become sensitive to MTH1 inhibitors, a genetic system was used where primary BJ cells are transformed in a step-wise transfection protocol using hTERT, SV40 large T antigen and Ras[22]. TH287 and TH588 were selectively cytotoxic to SV40-large-T- and Ras-expressing BJ cells (Fig. 4i and Extended Data Fig. 5o, q), demonstrating that MTH1 becomes important early on in the transformation process, when MTH1 expression also increases (Extended Data Fig. 7).

MutT rescues MTH1 inhibitor toxicity

Many small molecules are known to be promiscuous, for example several kinase inhibitors bind to multiple kinase targets. Because TH287 and TH588 are first-in-class inhibitors of the nudix hydrolase protein family, we wanted to determine how selective these compounds would be towards other members of the nudix protein family[23]. MTH2, NUDT5, NUDT12, NUDT14, NUDT16 and other proteins with known nucleoside triphosphate pyrophosphatase activity, that is, dCTPase, dUTPase and ITPA, were therefore expressed, purified and used to develop enzymatic activity assays. TH287 and TH588 were highly selective towards MTH1, with no relevant inhibition of any of the other tested proteins at 100 μM (Fig. 5a and Extended Data Fig. 8h, i). Although TH588 showed reasonable selectivity when tested on a much larger panel of 87 enzymes, GPCRs, kinases, ion channels and transporters at 10 μM (Extended Data Fig. 8g), it may still have other relevant off-target activities which are not yet known. In an effort to address this possibility, we purified the *Escherichia coli* MutT protein, which is the bacterial homologue of MTH1 (ref. 24) and has 8-oxodGTPase activity. We demonstrate that MutT is not inhibited by either TH287 or TH588 (Fig. 5b). Notably, MutT enzyme expressed in mitochondria or the nucleus (Extended Data Fig. 8k–m) partly rescues the cells from cytotoxicity and 53BP1 induction caused by TH287 and TH588 (Fig. 5c, d), demonstrating that restoration of

8-oxodGTPase activity in human cells reduces the toxicity of the compounds. The reason why MutT does not completely reverse compound toxicity is probably explained by its lack of 2-OH-dATPase activity[25], which is also contributing to toxicity.

In vivo activity of MTH1 inhibitors

Our data suggest that MTH1 inhibition may be a useful strategy to target cancer. Malignant melanomas are heterogeneous, exhibit high numbers of somatic mutations[4] and are generally resistant to treatments. Metastases derived from a patient with *BRAF*^*V600E*-mutated melanoma, resistant to carboplatin, dacarbazine and vemurafenib, was serially transplanted twice in mice. Ten mice carrying this tumour were randomized to receive TH588 once daily or vehicle control treatment. All TH588-treated mice responded with a reduced tumour growth rate, even though pharmacokinetic studies had shown that the serum level of TH588 dropped below IC₅₀ concentrations during the 24-h treatment (Fig. 5e,f). In addition, SW480 colorectal and MCF7 breast tumour xenografts also responded to once-daily TH588 treatment (Extended Data Fig. 4), supporting the view that targeting MTH1 is potentially useful for a range of different tumours.

Discussion

Here, we observe that the MTH1 protein is required for efficient survival of cancer cells, whereas it is non-essential in untransformed cells. To explain this phenomenon, we propose a model where dysfunctional redox regulation and ROS in cancer cells lead to an oxidized dNTP pool, which in turn requires the MTH1 protein to prevent incorporation and DNA damage after replication (Fig. 5g). MTH1 is found to be overexpressed in many cancers[26] (Extended Data Fig. 7) and the MTH1 catalytic activity is also markedly increased in lung cancer tissue[26], which is an example of non-oncogene addiction and probably a survival response to prevent incorporation of oxidized dNTPs. Possible reasons why normal cells seem to be less dependent on MTH1 activity could be due to rigid redox regulation in normal cells as compared to cancer cells, and/or intact cell cycle checkpoints responding to ROS[27] (Fig. 5g). Our rather simple model can potentially be complicated by the fact that
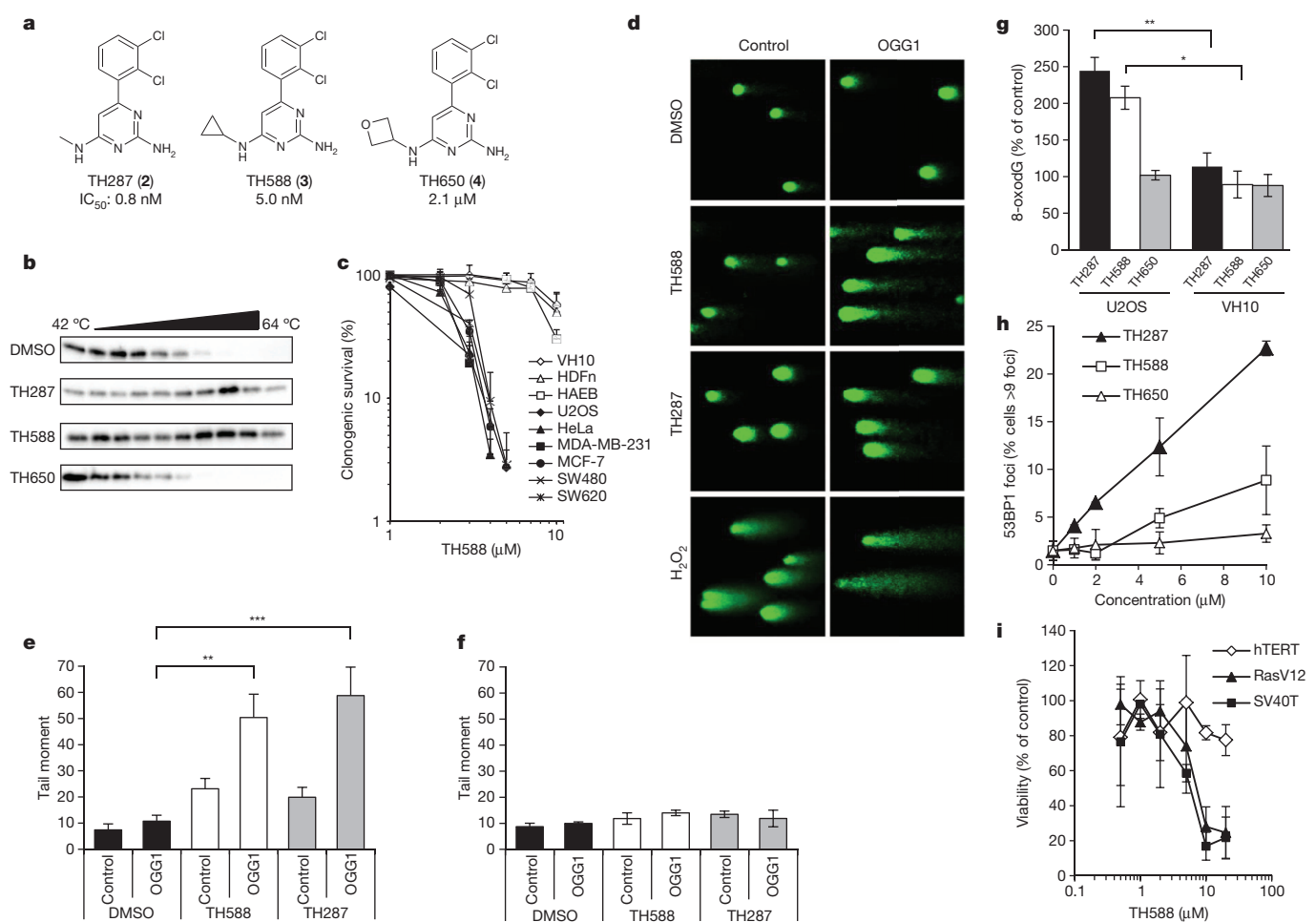
Figure 4 | Inhibition of MTH1 induces oxidative DNA damage and reduces survival in cancer cells. a, Chemical structure of the MTH1 inhibitors TH287 and TH588 and the poor inhibitor TH650. IC_{50} values shown as average of three independent experiments. **b**, Target engagement of MTH1 inhibitor to MTH1 protein in intact cells. The MTH1 protein precipitates in cells above 54 °C. Potent MTH1 inhibitors stabilize the MTH1 protein >54 °C. **c**, Clonogenic survival of primary/immortalized (open) or cancer cells (filled) after TH588 treatment. Values represent percentage of colonies in relation to DMSO-treated controls displayed as average ± s.d. from three independent experiments. **d–f**, U2OS cells (**d, e**) or VH10 cells (**f**) were treated with 10 μM MTH1 inhibitor for 24 h before being incubated with OGG1, to selectively cut out 8-oxodG from DNA, and run in alkaline Comet assay. H_2O_2 was used as control. Tail moment, calculated as % DNA in the tail multiplied by the tail length. Values represent average ± s.e.m. from three experiments; **$P < 0.01$, ***$P < 0.001$, one-way ANOVA. **g**, Quantification of 8-oxodGTP incorporation in DNA (avidin-AlexaFluor488 reactive substance) in U2OS and VH10 cells treated with MTH1 inhibitors (20 μM). Bars show average ± s.e.m. of three independent experiments. Asterisks mark a significant difference (*$P < 0.05$, **$P < 0.01$, one-sided Student's t-test). **h**, U2OS cells were seeded in 96-well plates and treated with MTH1 inhibitors for 72 h and 53BP1 foci formation determined. Values represent percentage of cells with >9 53BP1 foci per cell, average ± s.e.m. from three experiments. **i**, Viability of BJ cells with hTERT, hTERT/SV40T ('SV40T') or hTERT/SV40T/RasV12 ('RasV12') after 72 h TH588 treatment. Data shown as average ± s.e.m. from two independent experiments.

cancer cells may adapt to imbalanced redox status and activate powerful antioxidant mechanisms to allow survival, and therefore lower the overall ROS levels in cancer. Hence, re-balancing or further alterations of the redox status may hypothetically be a resistance mechanism to MTH1 inhibitors.

Previously, overexpressed MTH1 has been demonstrated to prevent Ras-induced DNA damage[12], and compounds selectively killing Ras-expressing cells are found, in a concurrent report, to work through targeting MTH1[28]. Here, we observe a similar cellular sensitivity to MTH1 inhibitors or siRNA in SV40 large T transfected BJ cells as those additionally transformed by Ras (Fig. 4i and Extended Data Figs 1 and 5), indicating that Ras expression is not a major determinant for the essential role of MTH1 in cancer.

MTH1 overexpression in mice promotes longevity[29], reduced anxiety[29] and protects from 3-nitropropionic-acid-induced neurodegeneration[30]. Upon pathological examination of ageing $Mth1^{-/-}$ mice a slight increase in cancer incidence is observed when grouping incidences in lung, liver and stomach cancers[13], whereas $Mth1$ gene disruption concomitantly

suppresses lung cancer development in $Ogg1^{-/-}$ mice[31]. These data suggest that long-term MTH1 inhibitor treatments may be associated with non-acute side effects, similar to those observed with PARP inhibitors. However, MTH1 inhibition may still be an interesting approach considering many anticancer target proteins are essential and severe side effects are associated with current chemotherapies.

Here we present first-in-class inhibitors to the nudix hydrolase protein family, which show excellent selectivity against other nudix family proteins and good selectivity towards kinases present in the selectivity panel. We expressed the bacterial MutT protein to rescue 8-oxodGTPase activity in cells in the presence of MTH1 inhibitors. We observed that MutT expression partly restores cellular viability and decreases DSB formation (53BP1 foci; Fig. 5). Proving that restoration of biochemical activity can rescue a biological effect is the ultimate proof for on-target effects of compounds and we suggest that such experiments should be used more widely.

In conclusion, we demonstrate that targeting MTH1 constitutes a novel cancer phenotypic lethal anticancer therapeutic approach targeting
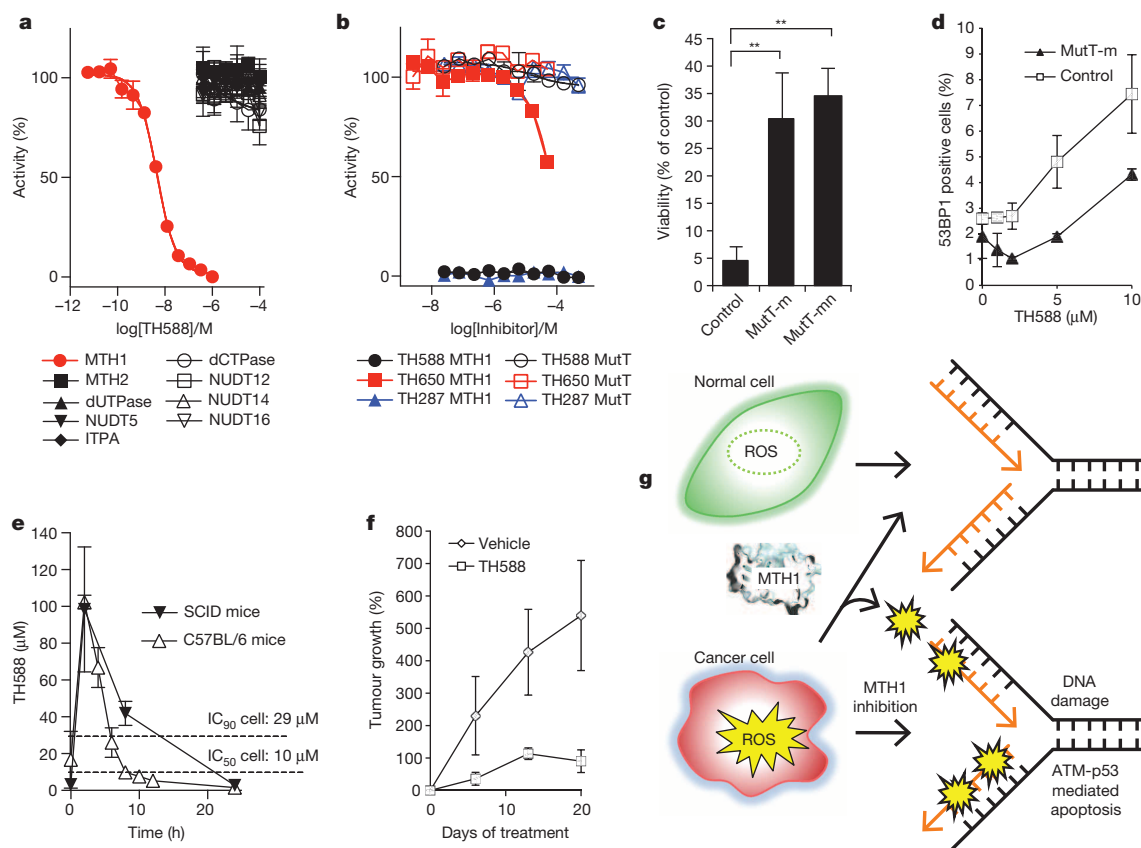
Figure 5 | TH588 reduces tumour growth via MTH1 inhibition. a, *In vitro* activity of human MTH1, MTH2 (NUDT15), NUDT5, NUDT12, NUDT14, NUDT16, dUTPase, dCTPase and ITPA treated with varying concentrations of the MTH1 inhibitor TH588. Data shown as average ± s.d. of three to six individual experiments, each performed in duplicate. **b**, MTH1 inhibitors TH287, TH588 and TH650 do not inhibit MutT. Inhibitory potency of TH287, TH588 and TH650 was tested against MutT (9.6 nM) and MTH1 (4.75 nM) at 100 μM dGTP and various inhibitor concentrations. Data shown as average ± s.d. from two independent experiments. **c, d**, Viability (**c**) and 53BP1 foci formation (**d**) in U2OS cells overexpressing MutT localized to the mitochondria (MutT-m) or MutT localized to both the mitochondria and nucleus (MutT-mn) or with no overexpression (control) treated with 10 μM

TH588 for 72 h. Data shown as average ± s.d. from three independent experiments, **P < 0.01, one-way ANOVA. Values represent average ± s.e.m. from three independent experiments. **e**, Plasma exposure of TH588 after 3 days treatment in C57BL/6 (n = 3 per time point) and SCID mice (n = 3 per time point) with subcutaneous injections of 25 mg kg⁻¹ TH588; plasma exposure was measured after the third injection on day 3. The dotted lines indicate the IC_{50} and IC_{90} values obtained from cell survival *in vitro* in SW480 cells. **f**, Growth of visceral metastasis from a malignant melanoma patient in NOD-scid IL2Rγ^null (NOG) mice treated once daily with TH588 (30 mg kg⁻¹, subcutaneous administration, once daily) or vehicle (n = 5 per group). **g**, Model for MTH1-dependent survival in cancer cells.

non-oncogene addiction, which converts deregulated metabolism and ROS levels present in cancer into toxic DNA damage.

METHODS SUMMARY

Cells were cultivated in suitable medium supplemented with 10% FBS at 37 °C at 5% CO_2. MTH1-puro vector was a gift from R. Weinberg. The siRNA-resistant MTH1 vector was made by site-directed mutagenesis of codons GAC and AGC at position 384–389 into GAT and TCC, coding for the same amino acids. The E56A mutant was generated by site-directed mutagenesis.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Druker, B. J. *et al.* Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nature Med.* **2**, 561–566 (1996).
2. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose)polymerase. *Nature* **434**, 913–917 (2005).
3. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
4. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
5. Zhang, Y. *et al.* Redox control of the survival of healthy and diseased cells. *Antioxid. Redox Signal.* **15**, 2867–2908 (2011).
6. Luo, M., He, H., Kelley, M. R. & Georgiadis, M. M. Redox regulation of DNA repair: implications for human health and cancer therapeutic development. *Antioxid. Redox Signal.* **12**, 1247–1269 (2010).
7. Topal, M. D. & Baker, M. S. DNA precursor pool: a significant target for N-methyl-N-nitrosourea in C3H/10T1/2 clone 8 cells. *Proc. Natl Acad. Sci. USA* **79**, 2211–2215 (1982).
8. Sakumi, K. *et al.* Cloning and expression of cDNA for a human enzyme that hydrolyzes 8-oxo-dGTP, a mutagenic substrate for DNA synthesis. *J. Biol. Chem.* **268**, 23524–23530 (1993).
9. Oka, S. *et al.* Two distinct pathways of cell death triggered by oxidative damage to nuclear and mitochondrial DNAs. *EMBO J.* **27**, 421–432 (2008).
10. Ichikawa, J. *et al.* Oxidation of mitochondrial deoxynucleotide pools by exposure to sodium nitroprusside induces cell death. *DNA Repair (Amst.)* **7**, 418–430 (2008).
11. Russo, M. T. *et al.* The oxidized deoxynucleoside triphosphate pool is a significant contributor to genetic instability in mismatch repair-deficient cells. *Mol. Cell. Biol.* **24**, 465–474 (2004).
12. Rai, P. *et al.* Enhanced elimination of oxidized guanine nucleotides inhibits oncogenic RAS-induced DNA damage and premature senescence. *Oncogene* **30**, 1489–1496 (2011).
13. Tsuzuki, T., Egashira, A. & Kura, S. Analysis of MTH1 gene function in mice with targeted mutagenesis. *Mutat. Res.* **477**, 71–78 (2001).
14. Schultz, L. B., Chehab, N. H., Malikzay, A. & Halazonetis, T. D. p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. *J. Cell Biol.* **151**, 1381–1390 (2000).
15. European Standards Committee on Oxidative DNA Damage. Measurement of DNA oxidation in human cells by chromatographic and enzymic methods. *Free Radic. Biol. Med.* **34**, 1089–1099 (2003).

16. Yoshimura, D. *et al.* An oxidized purine nucleoside triphosphatase, MTH1, suppresses cell death caused by oxidative stress. *J. Biol. Chem.* **278,** 37965–37973 (2003).
17. Narita, M. *et al.* Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell* **113,** 703–716 (2003).
18. Martinez Molina, D. *et al.* Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341,** 84–87 (2013).
19. Svensson, L. M. *et al.* Crystal structure of human MTH1 and the 8-oxo-dGMP product complex. *FEBS Lett.* **585,** 2617–2621 (2011).
20. Ushijima, Y. *et al.* A functional analysis of the DNA glycosylase activity of mouse MUTYH protein excising 2-hydroxyadenine opposite guanine in DNA. *Nucleic Acids Res.* **33,** 672–682 (2005).
21. German, P. *et al.* Activation of cellular signaling by 8-oxoguanine DNA glycosylase-1-initiated DNA base excision repair. *DNA Repair (Amst.)* **12,** 856–863 (2013).
22. Hahn, W. C. *et al.* Creation of human tumour cells with defined genetic elements. *Nature* **400,** 464–468 (1999).
23. McLennan, A. G. Substrate ambiguity among the nudix hydrolases: biologically significant, evolutionary remnant, or both? *Cell. Mol. Life Sci.* **70,** 373–385 (2013).
24. Maki, H. & Sekiguchi, M. MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature* **355,** 273–275 (1992).
25. Fujikawa, K. *et al.* The oxidized forms of dATP are substrates for the human MutT homologue, the hMTH1 protein. *J. Biol. Chem.* **274,** 18201–18205 (1999).
26. Speina, E. *et al.* Contribution of hMTH1 to the maintenance of 8-oxoguanine levels in lung DNA of non-small-cell lung cancer patients. *J. Natl Cancer Inst.* **97,** 384–395 (2005).
27. Guo, Z., Kozlov, S., Lavin, M. F., Person, M. D. & Paull, T. T. ATM activation by oxidative stress. *Science* **330,** 517–521 (2010).
28. Huber, K. V. M. *et al.* Stereospecific targeting of MTH1 by (*S*)-crizotinib as an anticancer strategy. *Nature* http://dx.doi.org/10.1038/nature13194 (this issue).
29. De Luca, G. *et al.* Prolonged lifespan with enhanced exploratory behavior in mice overexpressing the oxidized nucleoside triphosphatase hMTH1. *Aging Cell* **12,** 695–705 (2013).
30. De Luca, G. *et al.* A role for oxidized DNA precursors in Huntington's disease-like striatal neurodegeneration. *PLoS Genet.* **4,** e1000266 (2008).
31. Sakumi, K. *et al. Ogg1* knockout-associated lung tumorigenesis and its suppression by *Mth1* gene disruption. *Cancer Res.* **63,** 902–905 (2003).

Supplementary Information is available in the online version of the paper.

ARTICLE

Stereospecific targeting of MTH1 by (*S*)-crizotinib as an anticancer strategy

Kilian V. M. Huber[1], Eidarus Salah[2], Branka Radic[1], Manuela Gridling[1], Jonathan M. Elkins[2], Alexey Stukalov[1], Ann-Sofie Jemth[3], Camilla Göktürk[3], Kumar Sanjiv[3], Kia Strömberg[3], Therese Pham[3], Ulrika Warpman Berglund[3], Jacques Colinge[1], Keiryn L. Bennett[1], Joanna I. Loizou[1], Thomas Helleday[3], Stefan Knapp[2] & Giulio Superti-Furga[1]

Activated RAS GTPase signalling is a critical driver of oncogenic transformation and malignant disease. Cellular models of RAS-dependent cancers have been used to identify experimental small molecules, such as SCH51344, but their molecular mechanism of action remains generally unknown. Here, using a chemical proteomic approach, we identify the target of SCH51344 as the human mutT homologue MTH1 (also known as NUDT1), a nucleotide pool sanitizing enzyme. Loss-of-function of MTH1 impaired growth of KRAS tumour cells, whereas MTH1 overexpression mitigated sensitivity towards SCH51344. Searching for more drug-like inhibitors, we identified the kinase inhibitor crizotinib as a nanomolar suppressor of MTH1 activity. Surprisingly, the clinically used (*R*)-enantiomer of the drug was inactive, whereas the (*S*)-enantiomer selectively inhibited MTH1 catalytic activity. Enzymatic assays, chemical proteomic profiling, kinome-wide activity surveys and MTH1 co-crystal structures of both enantiomers provide a rationale for this remarkable stereospecificity. Disruption of nucleotide pool homeostasis via MTH1 inhibition by (*S*)-crizotinib induced an increase in DNA single-strand breaks, activated DNA repair in human colon carcinoma cells, and effectively suppressed tumour growth in animal models. Our results propose (*S*)-crizotinib as an attractive chemical entity for further pre-clinical evaluation, and small-molecule inhibitors of MTH1 in general as a promising novel class of anticancer agents.

Mutations in RAS isoforms are prevalent in human cancers, accompanied by poor prognosis and low survival, highlighting the need to develop new therapies[1–3]. Direct modulation of RAS activity has posed a significant challenge in drug discovery. Therefore, alternative approaches have been used, for example by interfering with RAS posttranslational modifications to prevent maturation and translocation of the active protein to the plasma membrane[4–6]. In addition, phenotypic screens have been used to search for small molecules that selectively target RAS-transformed cancer cells[7]. In 1995, this led to the discovery of a compound termed SCH51344 that suppressed the anchorage-independent growth of RAS-transformed fibroblasts[8]. As SCH51344 did not affect MAPK signalling, which is thought to be the primary mediator of RAS oncogenic activity, a novel but enigmatic mode of action was proposed[9].

Identification of MTH1 as the main target of SCH51344

We set out to identify the cellular targets of SCH51344 using a chemical proteomic strategy (Fig. 1a). We generated a SCH51344 affinity probe (Fig. 1b) which we incubated with lysates of KRAS-positive SW480 cells, which are sensitive to SCH51344, and analysed the binding proteins by mass spectrometry. High-affinity binders were discriminated against highly abundant low-affinity proteins by competition with the free unmodified compound. Bioinformatic analysis revealed the human 7,8-dihydro-8-oxoguanine triphosphatase MTH1 (also known as NUDT1) and adenosine kinase (ADK) as the primary cellular targets of SCH51344 (Fig. 1c). MTH1 has been implicated in aiding RAS-transformed cells to overcome oncogene-induced senescence by preventing reactive oxygen species (ROS)-induced DNA damage[10]. On the contrary, little was known about the role of ADK in malignant disease, but in line with published RNA interference data[11] we did not observe any growth impairment of SCH51344-sensitive[8] PANC1 human pancreatic carcinoma cells upon treatment with the ADK inhibitor

ABT-702 (data not shown). We therefore focused on MTH1 as the most likely relevant target of SCH51344. Having confirmed the binding of SCH51344 to MTH1 in both SW480 and DLD1 cells by immunoblot (Extended Data Fig. 1a), we used isothermal titration calorimetry (ITC) to determine a K_d value of 49 nM for SCH51344 (Fig. 1d and Extended Data Fig. 1b). MTH1 is a homologue of the bacterial mutT, a nucleotide pool sanitizing enzyme which cleaves oxidized nucleotides such as 8-oxo-deoxyguanosinetriphosphate (8-oxo-dGTP), thereby converting the triphosphates into the corresponding monophosphates[12]. The hydrolysis reaction ensures that the oxidized nucleotides can no longer be recognized by DNA polymerase, preventing the mispairing of bases during replication and thus transversion mutations[13,14]. To investigate the effect of SCH51344 on MTH1 catalytic activity we monitored the production of pyrophosphate (PPi) as a result of nucleotide triphosphate hydrolysis[15]. We determined half-maximum inhibitory concentration (IC$_{50}$) values of 215 nM, 410 nM and 675 nM for SCH51344 against the MTH1 substrates dGTP, 8-oxo-dGTP and 2-OH-dATP, respectively, confirming a direct effect of SCH51344 on MTH1 catalytic activity (Fig. 1e). To validate MTH1 as the causal target for the antiproliferative effects of SCH51344, we transfected human SW480 and DLD1 cells with *NUDT1* (MTH1) short interfering RNA that impaired colony formation (Fig. 1f). Stable knockdown using lentiviral short hairpin RNAs[16] phenocopied results obtained using the inhibitor (Extended Data Fig. 1c). Conversely, over-expression of MTH1[10] reduced sensitivity of SW480 cells against SCH51344 (Fig. 1g and Extended Data Fig. 1d), mechanistically corroborating the evidence that MTH1 is the main cellular target of SCH51344.

The (*S*)-crizotinib enantiomer inhibits MTH1 activity

Because SCH51344 has not been evaluated in a clinical setting, we decided to screen for other, more potent MTH1 inhibitors with favourable

---

[1]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria. [2]Nuffield Department of Clinical Medicine, Structural Genomics Consortium, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford OX3 7DQ, UK. [3]Science for Life Laboratory, Division of Translational Medicine and Chemical Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 17121 Stockholm, Sweden.
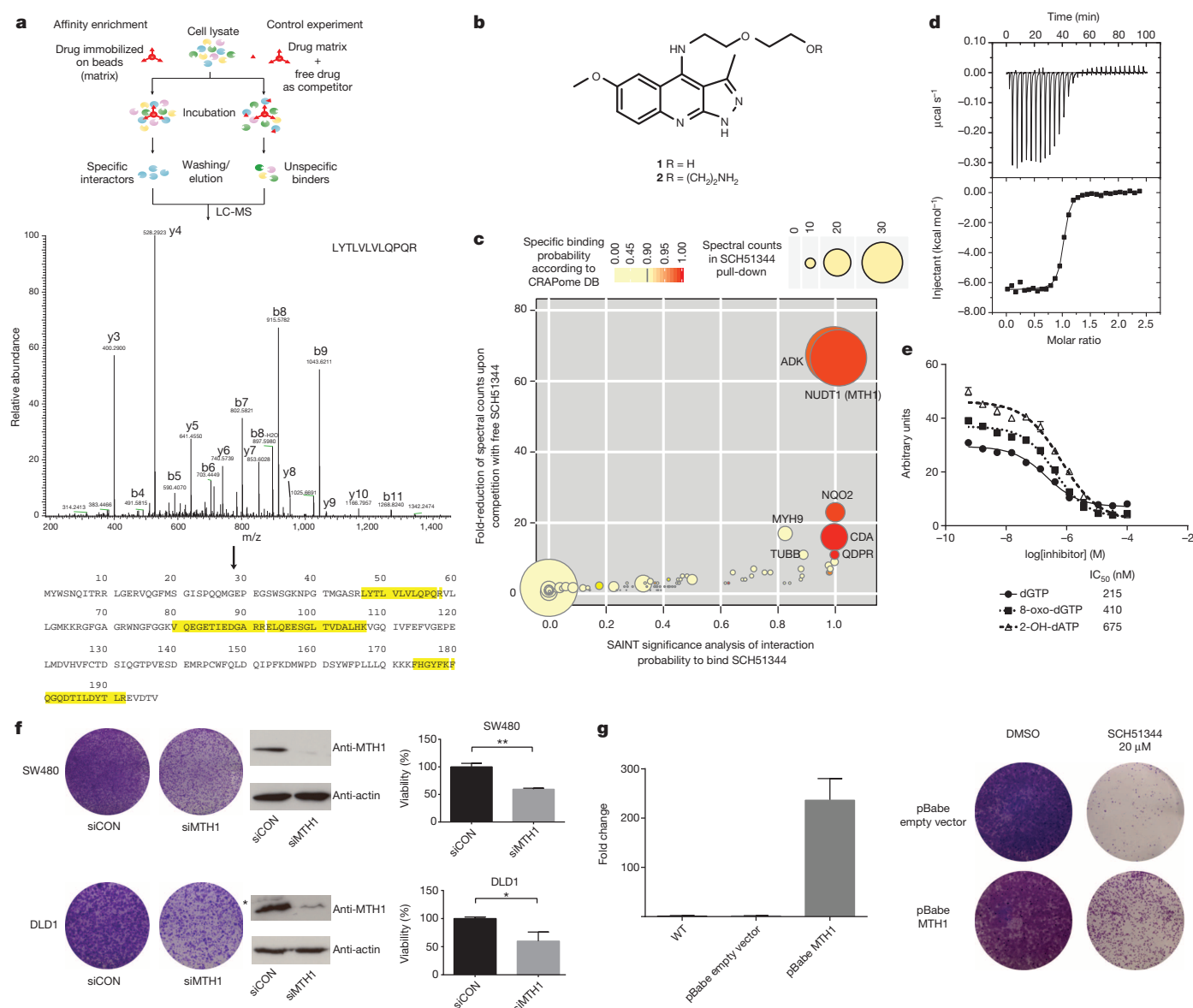
Figure 1 | MTH1 is the target of SCH51344. a, Representation of the chemical proteomic workflow. **b**, Structures of SCH51344 (**1**) and the probe used for affinity purification (**2**). **c**, Results from mass-spectrometry-based proteomic affinity purification experiment using SAINT and competition analysis. Data shown are based on two independent experiments for each condition ($n = 2$ per condition), and each replicate was analysed in two technical replicates. **d**, ITC data for MTH1 with SCH51344. The measured K_d was 49 nM ($n = 1$). **e**, SCH51344 inhibits hydrolysis of the MTH1 substrates dGTP, 8-oxo-dGTP and 2-OH-dATP, respectively. Data are shown for two technical replicates $\pm$ s.e.m. and are representative of at least duplicate experiments ($n \geq 2$). **f**, Silencing of MTH1 by siRNA impairs colony formation of KRAS-positive SW480 (top) and DLD1 (bottom) cells. Data shown as mean $\pm$ s.e.m. and images are representative of triplicate experiments ($n = 3$) ($P < 0.05$, t-test). Asterisk denotes unspecific band. **g**, MTH1 overexpression as monitored by real-time PCR (left) restores SW480 cell viability upon SCH51344 treatment (right). Data shown as mean $\pm$ s.e.m. and images are representative of three independent experiments ($n = 3$). WT, wild type.

pharmacokinetic and pharmacodynamic properties. On the basis of substrates and active site architecture we proposed that kinase inhibitors may target MTH1. Screening a kinase inhibitor collection in a thermal shift stability assay[17] we found that the dual c-MET/ALK inhibitor crizotinib[18,19] exhibited high affinity towards MTH1 (data not shown). Crizotinib recently received approval for the treatment of EML4-ALK-positive non-small cell lung cancer (NSCLC) and is in several other clinical trials[20–23]. However, using the catalytic MTH1 assay, we found that crizotinib batches obtained from different vendors resulted in varying IC$_{50}$ values. This could not be explained by impurities or degradation products as analytical data were in accordance with literature[18]. Because crizotinib bears a chiral centre, we speculated that variable amounts of crizotinib stereoisomers may occur in different batches of inhibitor. We prepared and tested both the pure, clinically used (R)- as well as the so

far unexplored (S)-enantiomer of crizotinib in the MTH1 catalytic assay, which suggested that the screening hit batch contained a racemic mixture. We found that pure (S)-crizotinib was a low nanomolar MTH1 inhibitor whereas the (R)-enantiomer gave IC$_{50}$ values in the micromolar range (Fig. 2a). These data were confirmed by direct-binding assays (ITC), indicating a 16-fold higher affinity of the (S)-enantiomer towards MTH1 (Fig. 2b and Extended Data Fig. 2a). Using K_m concentrations of substrates[12,15], we determined average IC$_{50}$ values of 330 nM and 408 nM for (S)-crizotinib and the MTH1 substrates 8-oxo-dGTP and 2-OH-dATP, respectively ($n = 2$). Consistent with these data, (S)-crizotinib efficiently inhibited colony formation of SW480 cells and KRAS-mutated PANC1 cells, similar to SCH51344 (Fig. 2c, d). In vitro K_d measurements indicated that (S)-crizotinib was considerably less potent than the (R)-enantiomer against the established targets ALK, MET and ROS1 (Extended
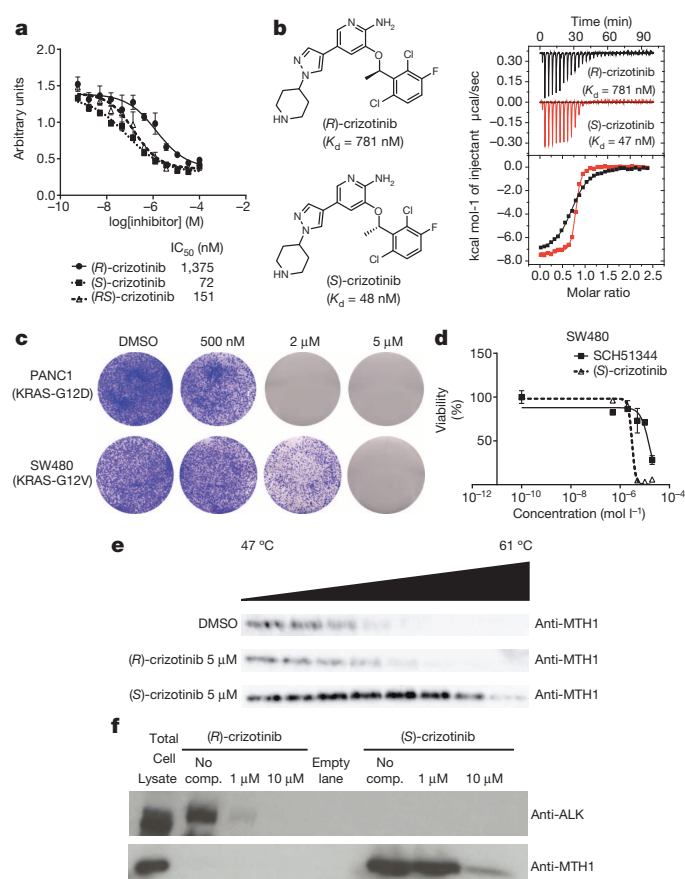
Figure 2 | (S)-Crizotinib is a nanomolar MTH1 inhibitor. a, MTH1 catalytic assay. Data are shown for both crizotinib enantiomers and the racemic mixture at 100 μM dGTP. Results indicate two technical replicates ± s.e.m. representative of at least duplicate experiments ($n \geq 2$). **b,** ITC for MTH1 with (R)- and (S)-crizotinib ($n = 1$). **c,** (S)-Crizotinib inhibits colony formation of PANC1 and SW480 cells. Images are representative of three independent experiments ($n = 3$). **d,** Comparison of antiproliferative efficacy of (S)-crizotinib versus SCH51344 against SW480 cells. Data shown as mean ± s.e.m. for three independent experiments ($n = 3$). **e,** Cellular thermal shift assay showing MTH1 target engagement by (S)-crizotinib in intact KRASV12-expressing BJ cells. Images are representative of two independent experiments ($n = 2$). **f,** The (S)-crizotinib affinity probe selectively binds MTH1, but not ALK, in SW480 lysates, whereas the (R)-enantiomer exerts inverse properties.

Data Fig. 2b). Treatment of SW480 cells with a specific c-MET inhibitor, a potential off-target for (S)-crizotinib[18], did not lead to the detection of any significant effects on proliferation (Extended Data Fig. 2c). However, investigating whether MTH1 overexpression could rescue SW480 cells from cell death induced by (S)-crizotinib in a similar manner as for SCH51344, we failed to observe any significant shift in IC_{50} values (Extended Data Fig. 2d), raising the question whether other targets contributed to the cell killing effect. We started investigating whether MTH1 was indeed targeted by (S)-crizotinib in intact cells. If a cellular protein is bound by a chemical agent, it is stabilized by the physical engagement compared to the non-engaged counterpart[24]. In a cellular thermal shift assay using BJ-KRASV12 cells, (S)-crizotinib, in contrast to (R)-crizotinib, efficiently stabilized MTH1 validating the differential targeting within cells (Fig. 2e).

Specificity of (S)-crizotinib and analysis of binding mode

To further investigate the ability of the two crizotinib enantiomers to engage cellular proteins, we derived chemical probes suitable for drug pull-downs (Supplementary Information). We tested two derivatized
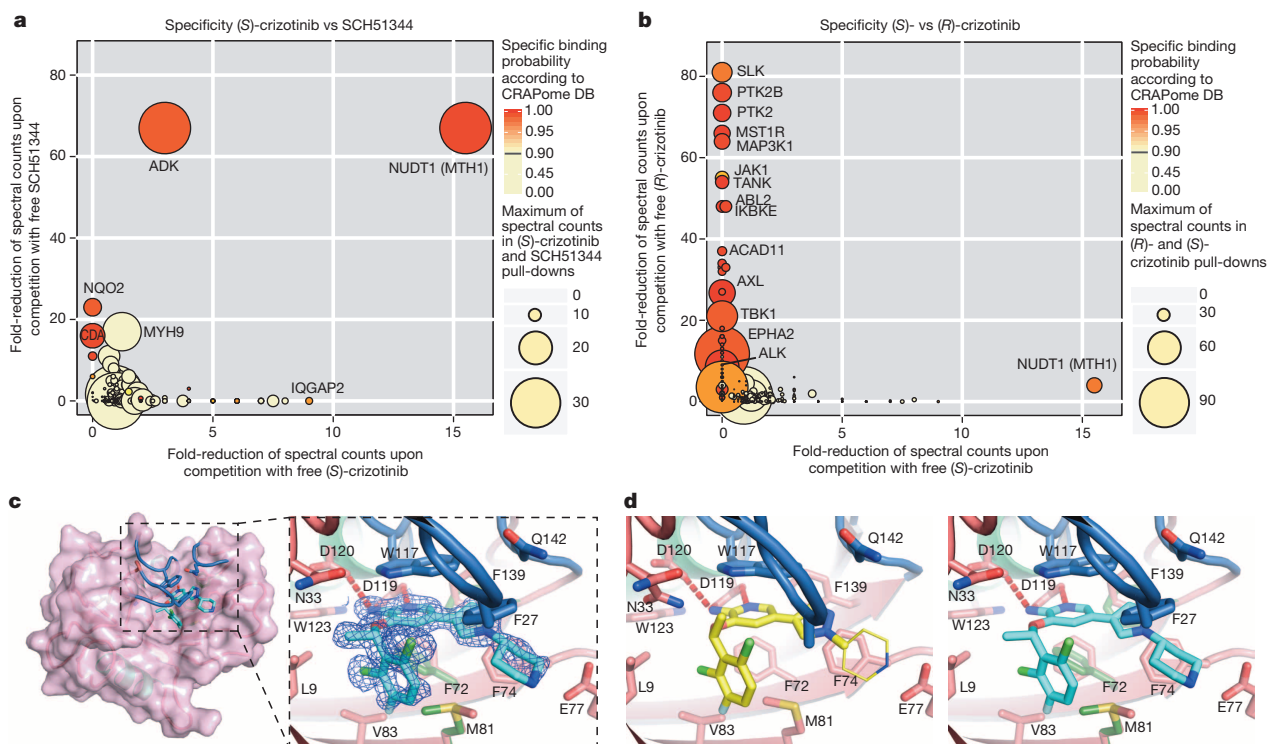
Figure 3 | Specificity and MTH1 co-crystal structure of (S)-crizotinib. a, b, Comparison of (S)-crizotinib specificity versus SCH51344 (**a**) and (R)-crizotinib (**b**). MTH1 is the only shared target with SCH51344 and is specific to (S)-crizotinib when compared to (R)-crizotinib. Data represent two independent experiments for each condition ($n = 2$ per condition), and each replicate was analysed in two technical replicates. **c,** Co-crystal structure of (S)-crizotinib and MTH1. MTH1 is in pink with light green alpha-helices and

the loops covering the binding site in blue. Hydrogen-bonding interactions are shown by dashed red lines. **d,** MTH1 interactions with (R)- and (S)-crizotinib. Left panel shows (R)-crizotinib in yellow; the thinner lines indicate part of the (R)-crizotinib that was not resolved in the electron density. Right panel shows (S)-crizotinib in cyan; alternate protein conformations in the absence of (S)-crizotinib are shown in dark green.

compounds for their ability to target ALK and MTH1 in SW480 cell extracts. The two enantiomers were remarkably specific for their cognate targets (Fig. 2f). If MTH1 was indeed the key target of (S)-crizotinib in other RAS-transformed cells, it should rank at the top of the specific interactors in an unbiased chemical proteomic experiment as done before with SCH51344. MTH1 was by far the most specific and prominent interactor of (S)-crizotinib (Extended Data Fig. 2e). Plotting the chemical proteomic results of SCH51344 and (S)-crizotinib against each other singled out MTH1 as the only common high-significance interactor (Fig. 3a). We also performed the reciprocal analysis with (R)-crizotinib which identified a plethora of protein kinases, all efficiently competed by free drug, but not MTH1 (Extended Data Fig. 2f). Notably, comparison of both profiles did not reveal any proteins that were significantly bound by both enantiomers (Fig. 3b). To exclude that either crizotinib enantiomer may target kinases of low abundance we interrogated a panel of 456 different recombinant kinases (KINOMEscan, Extended Data Fig. 3)[25]. In line with the chemoproteomic results the two enantiomers showed a remarkable stereoselectivity with very distinct profiles. The few kinases to which (S)-crizotinib showed some affinity were not calculated to be significantly inhibited. (R)-Crizotinib not only bound to ten times more kinases, but was also predicted to efficiently inhibit at least ten of them, including the well characterized cognate targets ALK and MET, but also LCK, IRAK1, JAK3, LOK (also known as STK10) and SLK. To understand the differences in MTH1 binding, we co-crystallized both (R)- and (S)-crizotinib with recombinant protein. The structure revealed an unfavourable eclipsed conformation of the methyl group at the chiral centre and a chlorine substituent on the benzyl ring is likely to reduce the energetic favourability of (R)-crizotinib binding

(Fig. 3c, d and Extended Data Figs 4 and 5). ITC data confirmed that the difference in binding between (R)- and (S)-crizotinib was entirely entropic and therefore not due to different binding interactions with the protein (Fig. 2b).

MTH1 inhibitors induce DNA damage in cancer cells

Because MTH1 is thought to prevent incorporation of oxidized nucleotides into DNA, we reasoned that our new MTH1 inhibitors should increase the content of genomic 8-oxo-guanine, and thus induce DNA damage. Immunofluorescence staining for both 53BP1 (also known as TP53BP1) and autophosphorylated ATM, specific markers for DNA damage, was increased in SW480 cells treated with MTH1 inhibitors (Fig. 4a and Extended Data Fig. 6a). 53BP1 foci, which we also observed in cells transfected with MTH1-siRNA, were enriched in nuclei of cells with higher levels of 8-oxo-guanine owing to increased genomic incorporation (Extended Data Fig. 6b, c). We also tried to quantify the oxidized nucleotides by high-performance liquid chromatography coupled with mass spectrometry; however, due to high experimental background, we failed to obtain reliable results. Because accumulation of 8-oxo-guanine should activate base-excision repair (BER)[16] and induce DNA single-strand breaks, we tested our inhibitors in an alkaline comet assay. Both (S)-crizotinib as well as SCH51344, but not (R)-crizotinib, yielded a significant tail moment, similar to cells transfected with MTH1-siRNA (Fig. 4b). Addition of the purified 8-oxo-guanine- or 2-hydroxy-adenine-specific DNA glycosylases, OGG1 and MUTYH, increased tail moments for (S)-crizotinib markedly, providing evidence for strong accumulation of these lesions upon inhibitor treatment (Extended Data Fig. 6d). MTH1 overexpression significantly reduced the number of DNA single-strand
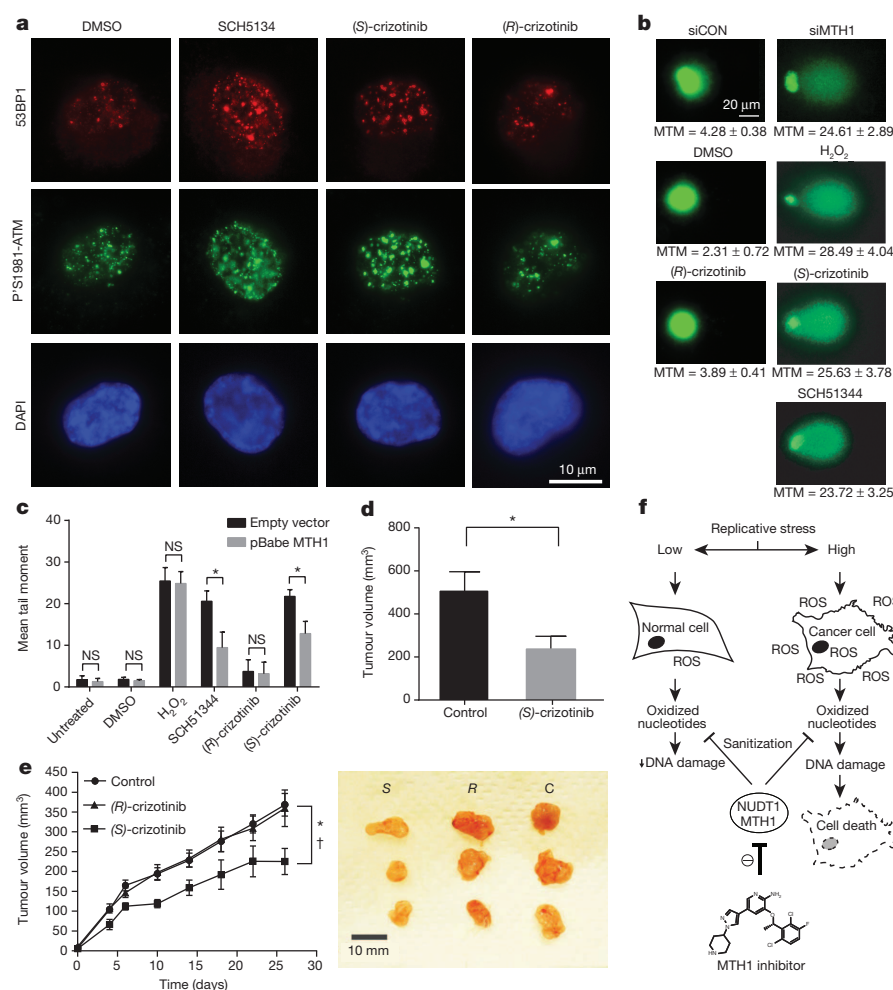
Figure 4 | (S)-Crizotinib is a selective MTH1 inhibitor with *in vivo* anticancer activity. a, The MTH1 inhibitors SCH51344 (5 μM) and (S)-crizotinib (2 μM), but not (R)-crizotinib (2 μM), induce DNA damage as indicated by an increase in 53BP1 foci and ATM autophosphorylation. Images are representative of three independent experiments (n = 3). **b**, Comet assay. Similar to MTH1 gene silencing both SCH51344 (5 μM) and (S)-crizotinib (2 μM), but not (R)-crizotinib (2 μM), induce DNA single-strand breaks. H$_2$O$_2$ was used as a positive control (150 μM, 10 min). Images are representative of three independent experiments (n = 3); data are shown as mean ± s.d. MTM, mean tail moment. **c**, MTH1 overexpression reduces the number of DNA single-strand breaks induced by SCH51344 and (S)-crizotinib. Compound concentrations are as in **b**. Data are shown as mean ± s.d. based on three independent experiments (n = 3). NS, not significant. **d**, Results from SW480 mouse xenograft study. Effect on tumour growth following 35-day treatment with the MTH1 inhibitor (S)-crizotinib (25 mg per kg, subcutaneously daily; data are shown as mean ± s.e.m., n = 8 per group). **e**, (S)-Crizotinib, but not (R)-crizotinib, impairs tumour growth in an SW480 colon carcinoma xenograft model (50 mg per kg, orally, daily). Data show mean ± s.e.m., n = 7–8 animals per group. Statistical analysis performed by two-way repeat measurement ANOVA, Sidak's multiple comparison; *P < 0.05 (S)-crizotinib vs control; †P < 0.05 (S)-crizotinib vs (R)-crizotinib. Images depict representative tumours for each treatment group (C, control). **f**, Proposed mechanism for MTH1-inhibitor-induced cancer cell death.

breaks induced by (S)-crizotinib as well as SCH51344, but not by H_2O_2 (Fig. 4c and Extended Data Fig. 6e), providing evidence for MTH1 being the functionally relevant target. To explore the role of p53 in the cellular response to MTH1 suppression[16] we created an SW480 Tet-on system[26] allowing for the inducible expression of p53 shRNAs and treated the cells with our inhibitors (Extended Data Fig. 7), which indicated a p53-independent mode of action. Treatment of SW480 cells expressing anti-MTH1 shRNA with the ATM- and ATR-inhibitors KU55933 and VE821, respectively, also did not display any differential effects (Extended Data Fig. 8). Similarly, $Atm^{-/-}$ mouse embryonic fibroblasts were equally sensitive to MTH1 inhibitors as their Atm-proficient counterparts. Investigating cell lines bearing additional mutations in DNA repair genes, we found that HCT116 cells deficient for p21 were particularly sensitive to (S)-crizotinib. We tested our inhibitors in BJ skin fibroblasts that were either wild type, immortalized by hTERT, or transformed by SV40T and/or mutant KRAS. Both SCH51344 and (S)-crizotinib showed highest toxicity towards the SV40T and KRASV12 cells (Extended Data Fig. 9). Importantly, when we treated wild type BJ cells with (R)- or (S)-crizotinib, we found that the (S)-enantiomer did not show any increased toxicity on non-transformed cells. Among a panel of human cancer cell lines, we consistently observed a strong antiproliferative effect for (S)-crizotinib, in line with its lower catalytic assay IC_{50} value. To explore the *in vivo* potential of (S)-crizotinib to impair tumour growth we performed mouse xenograft studies using SW480 cells. These experiments indicated that (S)-crizotinib, but not the (R)-enantiomer, was able to impair overall tumour progression as well as specifically reduce tumour volume by more than 50% (Fig. 4d, e and Extended Data Fig. 10a–c). This suggested that the two enantiomers have clearly diverse antitumour profiles and was consistent with their distinct molecular mechanism of action.

Targeting nucleotide pool homeostasis as cancer therapy

Increased levels of ROS in fast-proliferating cancer cells impair nucleotide pool homeostasis and contribute to mutations and DNA damage[10]. Removal of oxidized nucleotides by MTH1 may relieve cancer cells from proliferative stress and thereby represent a vulnerability factor and an attractive target for anticancer compounds (Fig. 4f)[27]. MTH1 levels are increased in RAS-expressing cancers (Extended Data Fig. 10d) ranging from lung cancer[28,29] to renal carcinoma[30], supportive of the notion that there is a connection between oncogenic transformation and oxidative stress. A potential more global role of MTH1 in tumorigenesis is supported by the observed antiproliferative effects for the inhibitors on cancer cells transformed by mechanisms other than RAS mutations (Extended Data Fig. 9c)[8]. Although prolonged clinical application will need to be evaluated critically in light of an increased long-term tumour burden in $Mth1^{-/-}$ (also known as $Nudt1^{-/-}$) mice, the mild phenotype of these knockout animals[31], and the specificity of MTH1 inhibitors speak for an appropriate therapeutic window. We propose that MTH1, together with other enzymes controlling sanitization of oxidized nucleotides, may represent a new attractive targeting strategy for difficult-to-treat tumours that display high levels of replicative and oxidative stress. The identification of the chemical mirror image of a recently clinically approved anticancer agent, crizotinib, as a nanomolar inhibitor of a yet pharmacologically unexploited cellular process, argues for further high-priority pre-clinical and clinical studies. A thorough investigation of the pharmacodynamics and pharmacokinetic properties of (S)-crizotinib will be necessary to understand why overexpression of MTH1 failed to rescue its cell-killing effects under the conditions tested. Until then, the possibility remains that targets other than MTH1 contribute to the effects of (S)-crizotinib. Whereas (S)-crizotinib is technically a new chemical entity and would require a new, separate drug approval process, the fact that it differs from a safe and bioavailable drug only in one chiral centre makes it somewhat more likely to have favourable, drug-like properties[32] and thus may be auspicious for an efficient evaluation of the potential therapeutic merits.

METHODS SUMMARY

Drug-affinity matrices were prepared by immobilizing 25 nmol of compound on 50 μl *N*-hydroxysuccinimide (NHS)-activated Sepharose 4 Fast Flow beads (GE Healthcare Bio-Sciences AB, Uppsala, Sweden). Detailed procedures for affinity chromatography, elution and mass spectrometry analyses are provided in the supporting information.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: weaving a tumorigenic web. *Nature Rev. Cancer* **11**, 761–774 (2011).
2. Parada, L. F., Tabin, C. J., Shih, C. & Weinberg, R. A. Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus *ras* gene. *Nature* **297**, 474–478 (1982).
3. Der, C. J., Krontiris, T. G. & Cooper, G. M. Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses. *Proc. Natl Acad. Sci. USA* **79**, 3637–3640 (1982).
4. Dekker, F. J. *et al.* Small-molecule inhibition of APT1 affects Ras localization and signaling. *Nature Chem. Biol.* **6**, 449–456 (2010).
5. Xu, J. *et al.* Inhibiting the palmitoylation/depalmitoylation cycle selectively reduces the growth of hematopoietic cells expressing oncogenic *Nras*. *Blood* **119**, 1032–1035 (2012).
6. Zimmermann, G. *et al.* Small molecule inhibition of the KRAS–PDEδ interaction impairs oncogenic KRAS signalling. *Nature* **497**, 638–642 (2013).
7. Yagoda, N. *et al.* RAS–RAF–MEK-dependent oxidative cell death involving voltage-dependent anion channels. *Nature* **447**, 865–869 (2007).
8. Kumar, C. C. *et al.* SCH 51344 inhibits *ras* transformation by a novel mechanism. *Cancer Res.* **55**, 5106–5117 (1995).
9. Walsh, A. B., Dhanasekaran, M., Bar-Sagi, D. & Kumar, C. C. SCH 51344-induced reversal of RAS-transformation is accompanied by the specific inhibition of the RAS and RAC-dependent cell morphology pathway. *Oncogene* **15**, 2553–2560 (1997).
10. Rai, P. *et al.* Enhanced elimination of oxidized guanine nucleotides inhibits oncogenic RAS-induced DNA damage and premature senescence. *Oncogene* **30**, 1489–1496 (2011).
11. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic *KRAS*-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
12. Fujikawa, K. *et al.* The oxidized forms of dATP are substrates for the human MutT homologue, the hMTH1 protein. *J. Biol. Chem.* **274**, 18201–18205 (1999).
13. Oka, S. *et al.* Two distinct pathways of cell death triggered by oxidative damage to nuclear and mitochondrial DNAs. *EMBO J.* **27**, 421–432 (2008).
14. Yoshimura, D. *et al.* An oxidized purine nucleoside triphosphatase, MTH1, suppresses cell death caused by oxidative stress. *J. Biol. Chem.* **278**, 37965–37973 (2003).
15. Svensson, L. M. *et al.* Crystal structure of human MTH1 and the 8-oxo-dGMP product complex. *FEBS Lett.* **585**, 2617–2621 (2011).
16. Rai, P. *et al.* Continuous elimination of oxidized nucleotides is necessary to prevent rapid onset of cellular senescence. *Proc. Natl Acad. Sci. USA* **106**, 169–174 (2009).
17. Fedorov, O. *et al.* Specific CLK inhibitors from a novel chemotype for regulation of alternative splicing. *Chem. Biol.* **18**, 67–76 (2011).
18. Cui, J. J. *et al.* Structure based drug design of crizotinib (PF-02341066), a potent and selective dual inhibitor of mesenchymal–epithelial transition factor (c-MET) kinase and anaplastic lymphoma kinase (ALK). *J. Med. Chem.* **54**, 6342–6363 (2011).
19. Zou, H. Y. *et al.* An orally available small-molecule inhibitor of c-Met, PF-2341066, exhibits cytoreductive antitumor efficacy through antiproliferative and antiangiogenic mechanisms. *Cancer Res.* **67**, 4408–4417 (2007).
20. Camidge, D. R. *et al.* Activity and safety of crizotinib in patients with ALK-positive non-small-cell lung cancer: updated results from a phase 1 study. *Lancet Oncol.* **13**, 1011–1019 (2012).
21. Kwak, E. L. *et al.* Anaplastic lymphoma kinase inhibition in non–small-cell lung cancer. *N. Engl. J. Med.* **363**, 1693–1703 (2010).
22. Gerber, D. E. & Minna, J. D. ALK inhibition for non-small cell lung cancer: from discovery to therapy in record time. *Cancer Cell* **18**, 548–551 (2010).
23. Butrynski, J. E. *et al.* Crizotinib in ALK-rearranged inflammatory myofibroblastic tumor. *N. Engl. J. Med.* **363**, 1727–1733 (2010).
24. Martinez Molina, D. *et al.* Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341**, 84–87 (2013).
25. Fabian, M. A. *et al.* A small molecule-kinase interaction map for clinical kinase inhibitors. *Nature Biotechnol.* **23**, 329–336 (2005).
26. Zuber, J. *et al.* An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance. *Genes Dev.* **25**, 1628–1640 (2011).
27. Sakumi, K. *et al.* Ogg1 knockout-associated lung tumorigenesis and its suppression by *Mth1* gene disruption. *Cancer Res.* **63**, 902–905 (2003).
28. Speina, E. *et al.* Contribution of hMTH1 to the maintenance of 8-oxoguanine levels in lung DNA of non-small-cell lung cancer patients. *J. Natl Cancer Inst.* **97**, 384–395 (2005).

29. Kennedy, C. H., Cueto, R., Belinsky, S. A., Lechner, J. F. & Pryor, W. A. Overexpression of *hMTH1* mRNA: a molecular marker of oxidative stress in lung cancer cells. *FEBS Lett.* **429,** 17–20 (1998).

30. Okamoto, K. *et al.* Overexpression of human *mutT* homologue gene messenger RNA in renal-cell carcinoma: evidence of persistent oxidative stress in cancer. *Int. J. Cancer* **65,** 437–441 (1996).

31. Tsuzuki, T. *et al.* Spontaneous tumorigenesis in mice defective in the *MTH1* gene encoding 8-oxo-dGTPase. *Proc. Natl Acad. Sci. USA* **98,** 11456–11461 (2001).

32. Hutt, A. J. Chirality and pharmacokinetics: an area of neglected dimensionality? *Drug Metabol. Drug Interact.* **22,** 79–112 (2007).

Supplementary Information is available in the online version of the paper.

ARTICLE

Structure of the LH1–RC complex from *Thermochromatium tepidum* at 3.0 Å

Satomi Niwa[1]*, Long-Jiang Yu[2]*, Kazuki Takeda[1]*, Yu Hirano[1], Tomoaki Kawakami[2], Zheng-Yu Wang-Otomo[2] & Kunio Miki[1]

The light-harvesting core antenna (LH1) and the reaction centre (RC) of purple photosynthetic bacteria form a supra-molecular complex (LH1–RC) to use sunlight energy in a highly efficient manner. Here we report the first near-atomic structure, to our knowledge, of a LH1–RC complex, namely that of a Ca^{2+}-bound complex from *Thermochromatium tepidum*, which reveals detailed information on the arrangement and interactions of the protein subunits and the cofactors. The RC is surrounded by 16 heterodimers of the LH1 αβ-subunit that form a completely closed structure. The Ca^{2+} ions are located at the periplasmic side of LH1. Thirty-two bacteriochlorophyll and 16 spirilloxanthin molecules in the LH1 ring form an elliptical assembly. The geometries of the pigment assembly involved in the absorption characteristics of the bacteriochlorophyll in LH1 and excitation energy transfer among the pigments are reported. In addition, possible ubiquinone channels in the closed LH1 complex are proposed based on the atomic structure.

The photosynthetic light reaction is a process by which light energy is converted into chemical energy. The primary process of the light reaction is the absorption of sunlight by a set of light-harvesting antenna protein complexes embedded in the membrane. When light is absorbed by the pigments in the antenna complexes, the energy is transferred rapidly and efficiently to the reaction centre (RC), where primary charge separation takes place across the membrane and a cyclic electron transport chain is formed. In purple photosynthetic bacteria, the antenna system contains mainly two types of light-harvesting complexes, LH1 and LH2. Both LH1 and LH2 are large oligomers of a basic structural subunit composed of a heterodimer of α- and β-apoproteins bound to light-absorbing pigments such as bacteriochlorophyll (BChl) and carotenoid. The atomic structure of the LH1–RC, including the arrangement of cofactors, is indispensable to the understanding of energy and electron transfer mechanisms.

Many structures of photosynthetic pigment–protein complexes have been hitherto determined, including a number of high-resolution structures of the RCs[1–3] and the LH2 complexes[4–6]. Nevertheless, only low-resolution structures of the LH1–RC have been reported. The structures of LH1 and LH1–RC from several species observed by cryo-electron microscopy and atomic force microscopy (AFM) exhibit a closed ring formed by 16 LH1 αβ-heterodimers[7–10], whereas the LH1 αβ-subunits of LH1–RC from *Rhodobacter* species show an S-shaped arrangement around the RC and an additional protein, PufX, seems to be involved in the assembly of the dimeric LH1–RC core complex[11–14]. Moreover, the crystal structure of the LH1–RC complex from *Rhodopseudomonas palustris* determined at 4.8 Å resolution[15] revealed an incomplete ring assembly composed of 15 heterodimers of LH1 αβ-apoproteins around the RC, and the electron density near the gap of the ring was assigned to a PufX-like polypeptide (protein W). This structure was verified by single-molecule spectroscopy[16]. However, owing to its low resolution this model has limited usefulness for elucidating the detailed arrangement of the subunits and cofactors.

Recently, we succeeded in obtaining high-quality crystals of the LH1–RC core complex from the thermophilic bacterium *Thermochromatium tepidum*[17]. This bacterium has an optimum growth temperature of approximately 50 °C and was discovered in the Mammoth Hot Springs of

Yellowstone National Park in Wyoming, USA[18], where a high concentration of calcium ions has been reported[19]. The calcium ions have been demonstrated to regulate the enhanced thermostability of the *T. tepidum* LH1–RC complex and the long wavelength of absorption maximum (Q_y) at 915 nm for the BChl *a* molecules (B915) in LH1[20–22]. The absorption at 915 nm is shifted to 880 nm by removal of calcium ions from the LH1–RC using EDTA, and the shift can be recovered by addition of calcium ions[20]. We report here the X-ray structure of *T. tepidum* LH1–RC at 3.0 Å resolution. Implications for the excitation energy transfer from B915 to the 'special-pair' BChls, the Ca^{2+}-regulation for the LH1 Q_y transition, and the ubiquinone shuttling through the closed LH1 ring, are discussed on the basis of the experimental structure.

Structure determination

Six native data sets at 3.0–3.5 Å resolutions were collected from the diffraction experiments. Crystals of four data sets belonged to the space group $P2_1$, whereas those of the remaining two belonged to the space group $C2$ (Extended Data Fig. 1 and Extended Data Table 1). Diffraction patterns for all the data sets showed significant anisotropy in intensities. The phases were calculated by the multiple isomorphous replacement method with anomalous scattering (MIRAS) assisted by the molecular replacement (MR) method (see Methods and Extended Data Table 2). Correction of the diffraction anisotropy and multi-crystal averaging significantly improved the electron density maps, with the densities for the side chains of the polypeptides becoming clearly observable (Fig. 1a and Extended Data Fig. 2a, b). In addition, the densities for cofactors such as BChl (Extended Data Fig. 2c), bacteriopheophytin (BPheo) and spirilloxanthin (Spx) were clearly observed.

Overall structure

The *T. tepidum* LH1–RC core complex is composed of cytochrome (Cyt), H, L and M subunits for the RC, $\alpha_{16}\beta_{16}$ subunits for the LH1 and 80 cofactors (Fig. 1b, c). The total molecular mass is ~380,000 Da. The RC is completely surrounded by the LH1 ring of 16 αβ-subunits. The LH1 forms an elliptical double-ring (Fig. 1c). The lengths of its major and minor axes are 82 Å and 73 Å for the inner ring and 105 Å and 96 Å for the outer ring, respectively. The structure of the RC with re-verified

[1]Department of Chemistry, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan. [2]Faculty of Science, Ibaraki University, Mito, Ibaraki 310-8512, Japan.
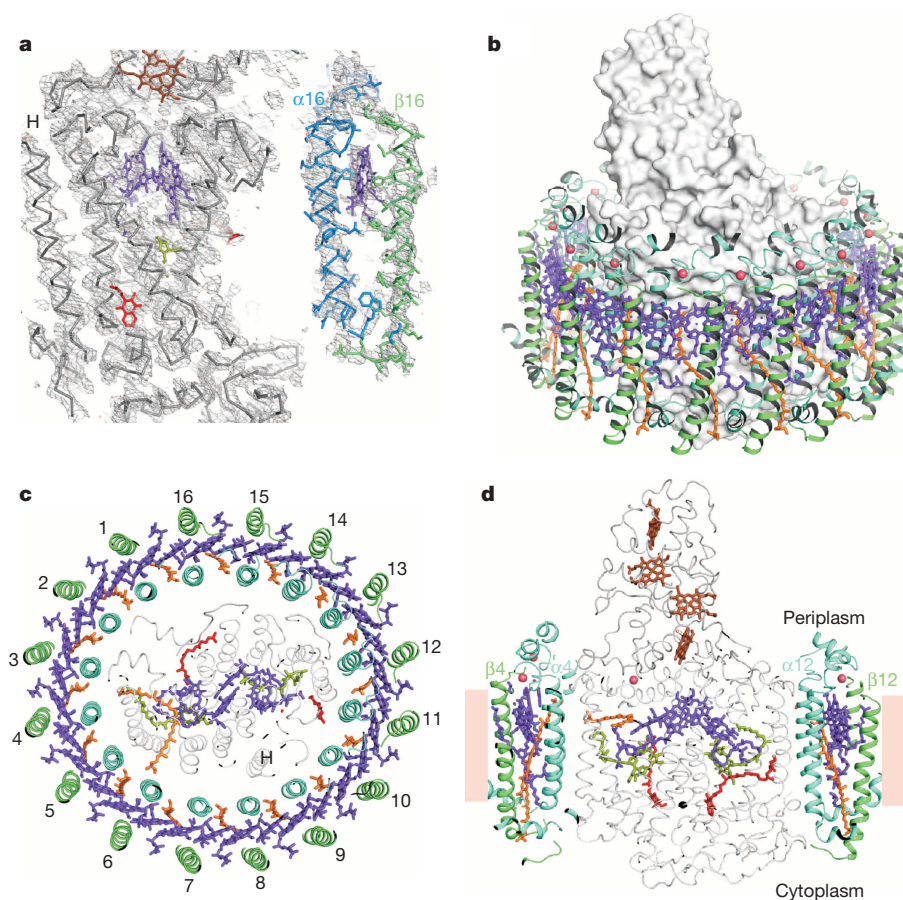*These authors contributed equally to this work.

Figure 1 | Architecture of the LH1–RC complex from *T. tepidum*. a, The experimental electron density map after density modification is represented as a grey mesh at a contour level of 1.0σ. The polypeptide chains and cofactors of the final model are also shown as Cα and stick models, in which the α16- and β16-apoproteins of LH1 are shown in blue and green, respectively. The H, L, M and Cyt subunits (grey) of the RC are also presented. In all panels, the following colour scheme is used for the cofactors. Haems, brown; calcium ions, pink; BChls, purple; BPheos, yellow green; Spx, orange; quinones, red; non-haem iron, black. **b**, Overall structure of the complex. The α and β-apoproteins in the LH1 ring are represented as blue and green ribbons, respectively. The surface of the RC portion is shown in grey. **c**, Arrangement of the cofactors is shown as a top view from the periplasmic side. **d**, A side view rotated by 90° from **c** around the horizontal axis. The lipid bilayer is schematically represented as a beige continuum.

amino acid sequences (Extended Data Fig. 3a) is very similar to that of the isolated RC (Protein Data Bank ID 1EYS)[3] (Extended Data Fig. 3b, c). However, a loop in the Cyt subunit (residues 172–196) shows a large deviation. A loop region in the H subunit (residues 44–58), which was invisible in the structure of isolated RC, could be constructed for the LH1–RC complex. The secondary quinone molecule (Q_B) in the proximity of the non-haem iron is also observed (Extended Data Fig. 2d), although it was not found in the isolated RC. In addition, a calcium ion at the interface of the Cyt and M subunits is identified (Extended Data Fig. 4). The LH1 ring is formed by the transmembrane helices of the inner α and outer β-apoproteins (Fig. 1d and Extended Data Fig. 5a, b). The amino- and carboxy-terminal ends of the α- and β-apoproteins are located on the cytoplasmic and periplasmic sides, respectively (Extended Data Fig. 5c, d). Both ends of the α-apoprotein form additional helices. The cytoplasmic helix is located outside the β ring. All 16 αβ-heterodimers can be superimposed completely over the transmembrane region (Extended Data Fig. 6a). However, the N-terminal 3_{10} helix and C-terminal helix of the α-apoprotein show relatively large deviations. This may be due to the intrinsic fluctuation of the region (Extended Data Fig. 6b, c). The N-terminal 3_{10} helix in particular takes diverse orientations relative to the transmembrane helices.

Each αβ-heterodimer subunit has two B915 molecules, one Spx, and one calcium ion. The stoichiometry is consistent with the biochemical analyses[17,20]. The pigment molecules are located in the middle of the α- and β-apoprotein rings. The two B915 molecules are liganded by α-His 36 and β-His 36, respectively. The α-B915 and β-B915 in a subunit

form a partially overlapping dimer with a face-to-face configuration (Fig. 1c).

Comparison with other antenna complexes

The structure of transmembrane regions of the LH1 αβ-apoproteins is similar to those of the LH2s from *Rhodopseudomonas acidophila* (Protein Data Bank ID 1NKZ)[4,23] and *Rhodospirillum molischianum* (Protein Data Bank ID 1LGH)[5] and the low-light LH2 (LH3) from *Rhodops. acidophila* (Protein Data Bank ID 1IJD)[6] (Extended Data Fig. 7a–f). In particular, *Rhodosp. molischianum* LH2 is very similar to the *T. tepidum* LH1 studied here, including with respect to the structure outside the transmembrane regions. Owing to the reduced curvature of the LH1 ring, the distance between the two transmembrane helices of the *T. tepidum* LH1 α-apoprotein is longer than those of LH2s, whereas that between the β-apoproteins is shorter (Extended Data Fig. 7g).

In the *Rhodops. palustris* LH1–RC model previously reported[15], LH1 is composed of 15 pairs of transmembrane helices of α- and β-apoproteins and BChls. The *Rhodops. palustris* LH1 ring has a gap interrupted by a transmembrane helix (protein W). This gap is proposed to be a gate through which the reduced ubiquinone can be transferred from the RC into the quinone pool outside the LH1 complex. Therefore, the putative protein W is thought to be a homologue of the PufX protein, which has a key role in the transfer of reduced ubiquinone in the *Rhodobacter* species. On the other hand, the *T. tepidum* LH1 has a closed ring structure, although it has approximately the same size and shape as *Rhodops. palustris* LH1–RC, excluding the Cyt subunit (Extended Data Fig. 7h).

The major difference from *Rhodops. palustris* LH1–RC is caused by the absence of a helix corresponding to the protein W. The gap is occupied by the 16th αβ-heterodimer in the *T. tepidum* LH1–RC (Fig. 1a). No *pufX*-like gene has been identified in the genome of *T. tepidum*[17,24].

Interactions between the proteins

The interaction in the transmembrane region between two neighbouring LH1 αβ-heterodimers is mainly mediated by the B915 and Spx molecules. Direct interactions between the transmembrane helices of the LH1 ring are exclusively observed at α-Phe 27 and α-Ile 29. These amino acids with large side-chains are frequently found in the LH1 α-apoproteins of purple bacteria (Extended Data Fig. 5a), whereas the corresponding amino acids in the LH2s are replaced with smaller residues. These residues may serve as spacers in forming the larger ring. Another pair, α-Gln 28 and β-Trp 28, connects the α- and β-apoproteins in the heterodimer subunit (Fig. 2a). There is no direct interaction between β-apoproteins.

The N-terminal domain on the cytoplasmic side of LH1 has fewer interactions between the α- and β-apoproteins. No direct bonds are formed between the α-apoproteins, or between the β-apoproteins. On the other hand, the C-terminal domain of LH1 on the periplasmic side undergoes several interactions, as indicated in Fig. 2b. Each α-chain forms hydrogen bonds with its neighbouring α-apoprotein through the side chains of α-Thr 42 and α-Asp 48, α-Asp 43 and the main chain carbonyl group of α-Leu 47. There are no direct interactions among the β-apoproteins in the C-terminal region. Side chains of the α-Asn 50 and β-Arg 43 in an αβ-heterodimer subunit form a hydrogen bond. The interactions between the *n*th α-apoprotein and (*n* + 1)th β-apoprotein are mediated by metal ions.

The metal ions were assigned to Ca^{2+} according to the anomalous difference Fourier maps calculated using diffraction data collected at long-wavelengths near the K-edge of the calcium (Extended Data Fig. 4). Sixteen Ca^{2+}-binding sites were identified in the C-terminal regions of the LH1 apoproteins, in agreement with the results of the previous biochemical analysis[22]. Each Ca^{2+} in the LH1 is coordinated with five

oxygen atoms from the side chains of α-Asp 49 and α-Asn 50, the main chain carbonyl of α-Trp 46, and the C-terminal carboxyl group of β-Leu 46 in the adjacent subunit (Fig. 2c). Two water molecules may also be involved in the ligation to form a seven-coordination geometry as frequently observed for the calcium chelating, although their electron densities have not been clearly identified. Because the amino acids at positions 49 and 50 of the α-apoprotein are not conserved in other species (Extended Data Fig. 5a), these residues are thought to be specific for the binding of calcium ions. Association of the α- and β-apoproteins through the Ca^{2+} stabilizes the LH1 structure and is therefore considered to account for its enhanced thermostability.

The α-B915 forms a hydrogen bond with α-Trp 46 through the C3-acetyl group (Fig. 2a). Similarly, the β-B915 forms a hydrogen bond with β-Trp 45 at the C3-acetyl group. This structural feature agrees with the findings of the resonance Raman study, which revealed a C3-acetyl C=O stretching band at 1,637 cm^{-1} and indicated strong hydrogen bonding for the carbonyl oxygen[25]. In addition, the β-B915 interacts with α-Gln 28 and β-Trp 28, forming direct hydrogen bonds with the carbonyl oxygen of the phytyl ester group (Fig. 2a). The C13-keto groups of B915 do not form hydrogen bonds.

The periplasmic half of LH1 apoproteins undergoes more interactions with RC than the cytoplasmic half (Fig. 3). The α-apoproteins, α5 and α13 (Fig. 1c), are located along the major axis of the elliptical LH1 ring and interact with the RC transmembrane helices. In the cytoplasmic half, some interactions are mediated by the positively charged residues, resulting in repulsion between the RC and LH1. Similar interactions are observed between α-apoproteins. The positively charged residues may be stabilized by interactions with anions such as sulphate ions and phosphatidyl groups of lipids, although they were not visible in the structure.

Arrangement of cofactors

The B915 and Spx molecules constitute an elliptical array in the LH1 ring (Fig. 4a). Thirty-two B915 molecules form a tightly stacked ring in a manner similar to the association of the LH2s. The B915 ring and the special-pair BChls of RC are aligned on the same level in the transmembrane region. The average Mg–Mg distance within a LH1 subunit is 9.04 ± 0.02 Å, whereas the distance between adjacent subunits is 8.46 ± 0.04Å. The distances are the shortest among the reported bacterial light-harvesting complexes, that is, LH1, LH2 and LH3. For the *T. tepidum* LH1–RC of this study, the LH1 Q_y transition occurs at 915 nm, which is approximately 30 nm red-shift from that of other purple bacteria. The Ca^{2+}-binding to the LH1 apoproteins is considered to make an additional contribution to the large red-shift, because the binding network is positioned close to the B915 molecules and could modify the configuration of the coupled pigments through the nearby α-Trp 46 and β-Trp 45, whose side chains are hydrogen-bonded to the C3-acetyl oxygen atoms of B915.
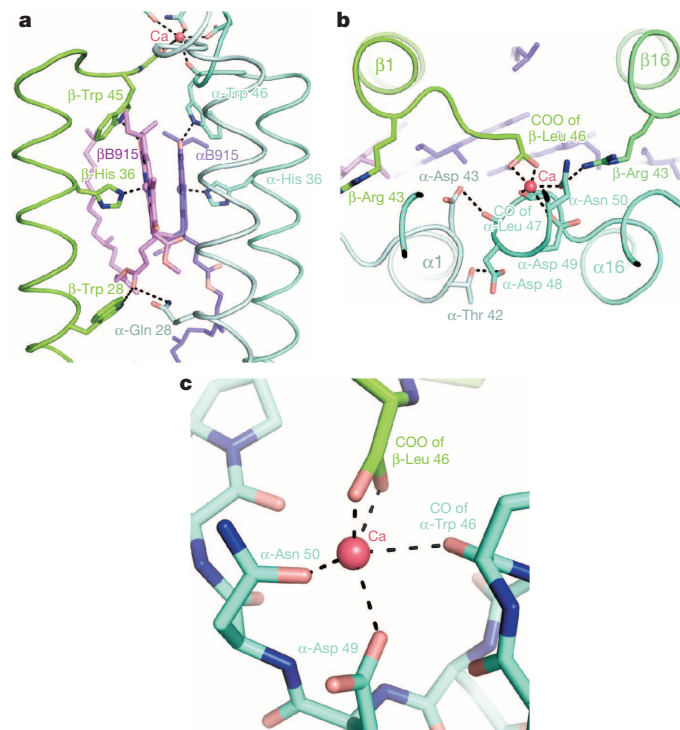

Figure 2 | Structural details of the LH1 ring. a, Interactions around the B915 BChl *a* molecules. **b,** Intra-LH1 ring interactions in the extra membrane region of the periplasmic side are shown as black dotted lines. **c,** Close-up view of the calcium-binding site.
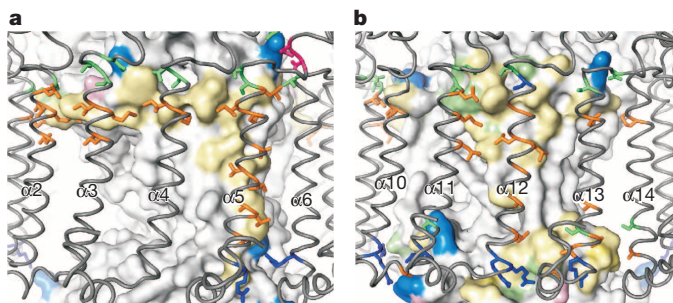

Figure 3 | Interactions between the LH1 and RC. a, Residues of the LH1 α-apoprotein involved in the interaction with the RC are coloured in orange (hydrophobic), green (neutral hydrophilic), blue (positively charged) and red (negatively charged). Contact areas on the molecular surface of the RC are coloured using the same scheme. **b,** A view rotated by 180° from **a** around the vertical axis.
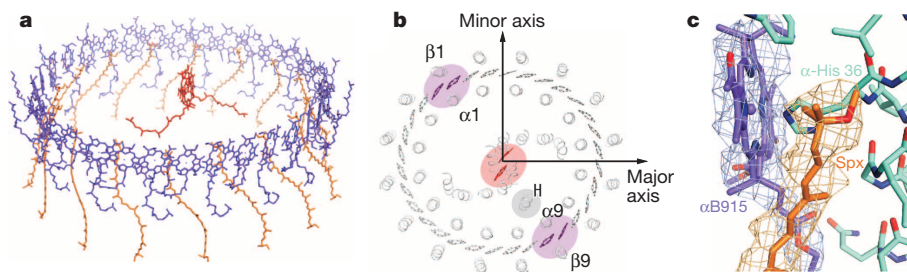
Figure 4 | The pigment arrangement. a, The supramolecular architecture of the pigments in LH1–RC. The B915 and Spx molecules are represented in purple and orange, respectively. Special-pair BChls are in red. **b**, Positions and orientations of the BChls in the LH1–RC complex are shown as a cross-section at the level of the special pair. The special pair and the B915 BChls with the closest orientations are coloured in red and purple, respectively. The transmembrane helix of H is shaded in grey. **c**, A close-up view shows the interaction between an αB915, Spx and the αB915-liganding His residue. Omit maps for the αB915 (light purple) and Spx (light orange) molecule are overlaid at contour levels of 3.0σ.

In the structure of *T. tepidum* LH1–RC, the α1-B915, β1-B915, α9-B915 and β9-B915 are almost parallel to the special-pair BChls (Fig. 4b). These LH1 pigments are positioned on a line that forms an angle of approximately 30° with the minor axis of the ellipse LH1 ring. The distances between the parallel B915s and special-pair BChls range 39.4–43.5 Å, whereas the shortest and the longest ones in the ring are approximately 37.4 Å and 49.0 Å, respectively. It is noted that the transmembrane helix of the H subunit in RC is located just between the αβ9-B915s and the special-pair BChls; and the N terminus of the L subunit, which is abundant in Trp residues, is located between the αβ1-B915s and the special-pair BChls.

Each all-*trans* Spx molecule is accommodated in the inter-helical space of an LH1 αβ-heterodimer subunit, and inclined at approximately 30° to the membrane normal. The Spx interacts with four transmembrane helices, an αβ-heterodimer, the neighbouring α-apoprotein and the opposite-side β-apoprotein. In addition, it interacts with the B915 dimer within a subunit and the αB915 in a neighbouring subunit. One of the methoxy groups of the Spx is in close proximity to the imidazole group of the α-His 36 that coordinates to the Mg atom of αB915 (Fig. 4c). The polyene chain of Spx is inclined at 30° to the direction of the Q_x transition moment of the B915 molecule. This geometry may explain the previously reported blue-shift of 7–10 nm for the LH1 Q_y transition caused by removal of the carotenoids[26].

Channels in LH1

The crystal structure provides insights into the mechanism of ubiquinone exchange in the LH1–RC core complex with a closed LH1 ring. This structure reveals that the LH1 ring possesses channels on the interface between each pair of adjacent αβ-heterodimers (Fig. 5a). The channels are located on the cytoplasmic side of the transmembrane region. The corresponding regions in LH2 are occupied by a set of monomeric BChls (B800), which are accommodated between the β-apoproteins and bound to the N-terminal helix of α-apoprotein. The Q_B site in the RC and the LH1 channels are almost on the same level in the transmembrane region (Fig. 5b). There is a large space between the Q_B site and the LH1 inner wall, as the closest distance between the Q_B site and the channel inner gate is approximately 30 Å.

The channels are not straight and have an average size approximately equal to that of the benzoquinone head of an ubiquinone. The dimension of the narrowest portion of the channel is 9×4 Å, whereas the width is larger than 12×6 Å at the entrance. The narrowest portion of the channels is defined by the LH1 α-apoproteins and pigment molecules. These include the phytyl groups of B915, Spxs and the side chains of α-Val 20, α-Val 22, α-Ile 24 and α-Val 25, which form a flexible and hydrophobic environment for the interior of the channels. There are variations of about 2 Å in the inter-helix distances for both the α and β rings (Extended Data Fig. 8a). As a result, the size and shape of the channel openings are variable (Extended Data Fig. 8b, c). This may reflect the intrinsic nature of the breathing motion of the LH1 complex, as speculated previously[27]. Studies using AFM and cryo-electron microscopy have shown that LH1 complexes in both the LH1-only and LH1–RC forms are quite flexible assemblies, with variations in shape from circles to ellipses[28,29]. The flexibility may be partly related to the lack of interactions between the adjacent αβ-heterodimer subunits on the cytoplasmic side, as shown above. From these properties of the channels, it is conceivable that the multiple channels could be available simultaneously for the ubiquinone shuttling through the closed LH1 ring. This notion is supported by the results of spectroscopic studies[30,31] and molecular dynamics simulation[32].

Conclusion

The near-atomic structure of the LH1–RC core complex reveals many details of the arrangement of all proteins and cofactors as well as the interactions among them. These results enable us to understand the molecular mechanism of the energy transfer from LH1 to RC, and the ubiquinone shuttling. Moreover, identification of the calcium-binding sites in the *T. tepidum* LH1–RC provides a structural basis for understanding the unique properties of its unusual LH1 Q_y transition. This study will provide invaluable structural information and stimulate further spectroscopic studies and simulation analyses on the mechanism of excitation energy and electron transfers in photosynthesis.

METHODS SUMMARY

LH1–RC was extracted from cells of *T. tepidum* and purified with *n*-decylphosphocholine (DPC). The *T. tepidum* LH1–RC was crystallized into two space groups, $P2_1$ ($a = 167.2$ Å, $b = 145.4$ Å, $c = 210.5$ Å, $\beta = 108.5°$) and C2 ($a = 227.3$ Å, $b = 148.2$ Å, $c = 161.8$ Å, $\beta = 117.6°$), at 20 °C in a solution containing 22 mg ml^{-1} *T. tepidum* LH1–RC, 20 mM Tris-HCl (pH 7.4), 50 mM CaCl$_2$, 0.54% DPC, 2% glycerol and 14% PEG3000. The diffraction data sets were collected at SPring-8 and the Photon Factory in Japan. The structure was determined using the MIRAS method. The structure was refined to 3.0 Å resolution.
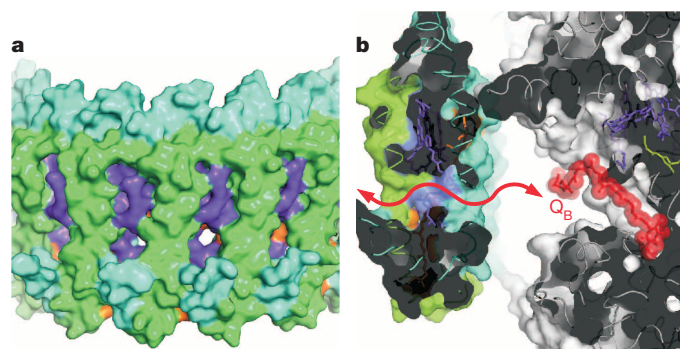

Figure 5 | Channels in the LH1 ring. a, The molecular surface of the LH1 ring is viewed from the outside. The surfaces of the α- and β-apoproteins, B915 and Spxs are coloured in blue, green, purple and orange, respectively. **b**, A cross section of the *T. tepidum* LH1–RC is shown for the interior of the channel. The red arrow illustrates a diffusion route for the ubiquinone.

1. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3Å resolution. *Nature* **318,** 618–624 (1985).
2. Allen, J. P., Feher, G., Yeates, T. O., Komiya, H. & Rees, D. C. Structure of the reaction center from *Rhodobacter sphaeroides* R-26: the protein subunits. *Proc. Natl Acad. Sci. USA* **84,** 6162–6166 (1987).
3. Nogi, T., Fathir, I., Kobayashi, M., Nozawa, T. & Miki, K. Crystal structures of photosynthetic reaction center and high-potential iron-sulfur protein from *Thermochromatium tepidum*: thermostability and electron transfer. *Proc. Natl Acad. Sci. USA* **97,** 13561–13566 (2000).
4. McDermott, G. *et al.* Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. *Nature* **374,** 517–521 (1995).
5. Koepke, J., Hu, X., Muenke, C., Schulten, K. & Michel, H. The crystal structure of the light-harvesting complex II (B800–850) from *Rhodospirillum molischianum*. *Structure* **4,** 581–597 (1996).
6. McLuskey, K., Prince, S. M., Cogdell, R. J. & Isaacs, N. W. The crystallographic structure of the B800-820 LH3 light-harvesting complex from the purple bacteria *Rhodopseudomonas acidophila* strain 7050. *Biochemistry* **40,** 8783–8789 (2001).
7. Karrasch, S., Bullough, P. A. & Ghosh, R. The 8.5 Å projection map of the light-harvesting complex I from *Rhodospirillum rubrum* reveals a ring composed of 16 subunits. *EMBO J.* **14,** 631–638 (1995).
8. Jamieson, S. J. *et al.* Projection structure of the photosynthetic reaction centre-antenna complex of *Rhodospirillum rubrum* at 8.5 Å resolution. *EMBO J.* **21,** 3927–3935 (2002).
9. Scheuring, S. *et al.* Nanodissection and high-resolution imaging of the *Rhodopseudomonas viridis* photosynthetic core complex in native membranes by AFM. *Proc. Natl Acad. Sci. USA* **100,** 1690–1693 (2003).
10. Scheuring, S. *et al.* Structural role of PufX in the dimerization of the photosynthetic core complex of *Rhodobacter sphaeroides*. *J. Biol. Chem.* **279,** 3620–3626 (2004).
11. Jungas, C., Ranck, J. L., Rigaud, J. L., Joliot, P. & Verméglio, A. Supramolecular organization of the photosynthetic apparatus of *Rhodobacter sphaeroides*. *EMBO J.* **18,** 534–542 (1999).
12. Siebert, C. A. *et al.* Molecular architecture of photosynthetic membranes in *Rhodobacter sphaeroides*: the role of PufX. *EMBO J.* **23,** 690–700 (2004).
13. Qian, P., Hunter, C. N. & Bullough, P. A. The 8.5 Å projection structure of the core RC-LH1-PufX dimer of *Rhodobacter sphaeroides*. *J. Mol. Biol.* **349,** 948–960 (2005).
14. Scheuring, S., Busselez, J. & Lévy, D. Structure of the dimeric PufX-containing core complex of *Rhodobacter blasticus* by *in situ* atomic force microscopy. *J. Biol. Chem.* **280,** 1426–1431 (2005).
15. Roszak, A. W. *et al.* Crystal structure of the RC-LH1 core complex from *Rhodopseudomonas palustris*. *Science* **302,** 1969–1972 (2003).
16. Richter, M. F. *et al.* Refinement of the X-ray structure of the RC-LH1 core complex from *Rhodopseudomonas palustris* by single-molecule spectroscopy. *Proc. Natl Acad. Sci. USA* **104,** 20280–20284 (2007).
17. Suzuki, H. *et al.* Purification, characterization and crystallization of the core complex from thermophilic purple sulfur bacterium *Thermochromatium tepidum*. *Biochim. Biophys. Acta* **1767,** 1057–1063 (2007).
18. Madigan, M. T. A novel photosynthetic purple bacterium isolated from a Yellowstone hot spring. *Science* **225,** 313–315 (1984).
19. Rowe, J. J., Fournier, R. O. & Morey, G. W. Chemical analysis of thermal waters in Yellowstone National Park, Wyoming, 1960–65. U.S. Geological Survey Bulletin 1303 http://pubs.usgs.gov/bul/1303/report.pdf (1973).
20. Kimura, Y. *et al.* Calcium ions are involved in the unusual red shift of the light-harvesting 1 Q_y transition of the core complex in thermophilic purple sulfur bacterium *Thermochromatium tepidum*. *J. Biol. Chem.* **283,** 13867–13873 (2008).
21. Kimura, Y., Yu, L. J., Hirano, Y., Suzuki, H. & Wang, Z.-Y. Calcium ions are required for the enhanced thermal stability of the light-harvesting-reaction center core complex from thermophilic purple sulfur bacterium *Thermochromatium tepidum*. *J. Biol. Chem.* **284,** 93–99 (2009).
22. Yu, L. J., Kato, S. & Wang, Z.-Y. Examination of the putative Ca²⁺-binding site in the light-harvesting complex 1 of thermophilic purple sulfur bacterium *Thermochromatium tepidum*. *Photosynth. Res.* **106,** 215–220 (2010).
23. Papiz, M. Z., Prince, S. M., Howard, T., Cogdell, R. J. & Isaacs, N. W. The structure and thermal motion of the B800–850 LH2 complex from *Rps. acidophila* at 2.0 Å resolution and 100 K: new structural features and functionally relevant motions. *J. Mol. Biol.* **326,** 1523–1538 (2003).
24. Fathir, I. *et al.* Biochemical and spectral characterization of the core light harvesting complex 1 (LH1) from the thermophilic purple sulfur bacterium *Chromatium tepidum*. *Photosynth. Res.* **58,** 193–202 (1998).
25. Kimura, Y. *et al.* Metal cations modulate the bacteriochlorophyll–protein interaction in the light-harvesting 1 core complex from *Thermochromatium tepidum*. *Biochim. Biophys. Acta* **1817,** 1022–1029 (2012).
26. Davis, C. M., Bustamante, P. L. & Loach, P. A. Reconstitution of the bacterial core light-harvesting complexes of *Rhodobacter sphaeroides* and *Rhodospirillum rubrum* with isolated α- and β-polypeptides, bacteriochlorophyll *a*, and carotenoid. *J. Biol. Chem.* **270,** 5793–5804 (1995).
27. Walz, T. & Ghosh, R. Two-dimensional crystallization of the light-harvesting I-reaction centre photounit from *Rhodospirillum rubrum*. *J. Mol. Biol.* **265,** 107–111 (1997).
28. Bahatyrova, S. *et al.* Flexibility and size heterogeneity of the LH1 light harvesting complex revealed by atomic force microscopy. *J. Biol. Chem.* **279,** 21327–21333 (2004).
29. Scheuring, S. in *The purple phototrophic bacteria* (eds Hunter, C. N., Daldal, F. & Beatty, J. T.) 941–952 (Springer, 2009).
30. Comayras, F., Jungas, C. & Lavergne, J. Functional consequences of the organization of the photosynthetic apparatus in *Rhodobacter sphaeroides*: II. A study of PufX⁻ membranes. *J. Biol. Chem.* **280,** 11214–11223 (2005).
31. Mascle-Allemand, C., Lavergne, J., Bernadac, A. & Sturgis, J. N. Organisation and function of the *Phaeospirillum molischianum* photosynthetic apparatus. *Biochim. Biophys. Acta* **1777,** 1552–1559 (2008).
32. Aird, A., Wrachtrup, J., Schulten, K. & Tietz, C. Possible pathway for ubiquinone shuttling in *Rhodospirillum rubrum* revealed by molecular dynamics simulation. *Biophys. J.* **92,** 23–33 (2007).

Author Contributions Z.-Y.W.-O. and K.M. initiated the project and supervised all experiments. L.-J.Y. carried out experiments of purification, crystallization and post-crystallization treatment. Y.H. conducted screenings of crystallization condition. L.-J.Y., Y.H. and T.K. collected X-ray diffraction data. S.N. and K.T. carried out structure determination. All authors contributed to the data analysis. S.N., K.T., L.-J.Y., Z.-Y.W.-O. and K.M. prepared the manuscript. All authors approved the final version of the manuscript.

Author Information The coordinates and structure factors for the *P*2₁ and *C*2 crystals have been deposited in the Protein Data Bank (http://www.pdb.org/pdb/home/home.do) under accession numbers 3WMN, 3WMO and 3WMM, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.M. (miki@kuchem.kyoto-u.ac.jp) or Z.-Y.W.-O. (wang@ml.ibaraki.ac.jp).

LETTER

Thermal fatigue as the origin of regolith on small asteroids

Marco Delbo[1], Guy Libourel[2,3], Justin Wilkerson[4], Naomi Murdoch[1,5], Patrick Michel[1], K. T. Ramesh[4], Clément Ganino[3], Chrystele Verati[3] & Simone Marchi[6]

Space missions[1,2] and thermal infrared observations[3] have shown that small asteroids (kilometre-sized or smaller) are covered by a layer of centimetre-sized or smaller particles, which constitute the regolith. Regolith generation has traditionally been attributed to the fall back of impact ejecta and by the break-up of boulders by micrometeoroid impact[4,5]. Laboratory experiments[6] and impact models[4], however, show that crater ejecta velocities are typically greater than several tens of centimetres per second, which corresponds to the gravitational escape velocity of kilometre-sized asteroids. Therefore, impact debris cannot be the main source of regolith on small asteroids[4]. Here we report that thermal fatigue[7–9], a mechanism of rock weathering and fragmentation with no subsequent ejection, is the dominant process governing regolith generation on small asteroids. We find that thermal fragmentation induced by the diurnal temperature variations breaks up rocks larger than a few centimetres more quickly than do micrometeoroid impacts. Because thermal fragmentation is independent of asteroid size, this process can also contribute to regolith production on larger asteroids. Production of fresh regolith originating in thermal fatigue fragmentation may be an important process for the rejuvenation of the surfaces of near-Earth asteroids, and may explain the observed lack of low-perihelion, carbonaceous, near-Earth asteroids[10].

The collisional and gravitational re-accumulation processes by which small asteroids are formed probably result in the creation of surfaces composed of boulders[11]. These boulders are broken up by micrometeoroid impacts into the smaller particles constituting the regolith[4,5]. A standard model[12] that calculates the time required to form regolith by fragmenting rocks of sizes between 1 and 10 cm by micrometeoroid impacts shows that these rocks on the Moon's surface will be broken down into smaller rocks in several million years[5,12]. Using the known[13] orbital distribution of micrometeoroids and a method[14] to calculate the impact probability of micrometeoroids with the Moon and asteroids (Methods), we find that the breakdown of surface rocks requires about the same amount of time on near-Earth asteroids (NEAs; asteroids with a perihelion distance of $q < 1.3$ AU, where 1 AU is the Earth–Sun distance) and on the Moon, whereas on main-belt asteroids (MBAs) this time is about ten times longer than on the Moon (Fig. 1).

Boulders on the surfaces of asteroids are also exposed to cyclic diurnal temperature variations, which cause mechanical stresses. To answer the question of whether these stresses are large enough to induce thermal
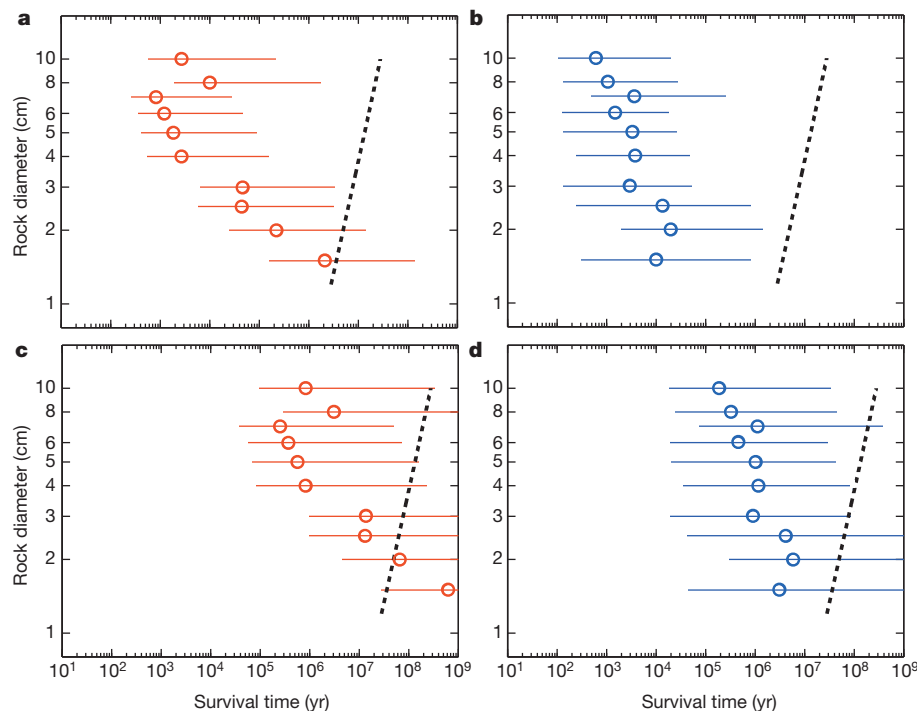

Figure 1 | Time required to break rocks on asteroids. Symbols show the time required to thermally fragment 90% of rocks for the nominal values of the model parameters. The thick dashed lines show the times at which 90% of these same rocks are broken by micrometeoroid impacts. **a**, Ordinary chondrite-like asteroid 1 AU from the Sun; **b**, carbonaceous chondrite-like asteroid at 1 AU; **c**, ordinary chondrite-like asteroid at 2.5 AU; **d**, carbonaceous chondrite-like asteroid at 2.5 AU. Error bars show the change in the thermal fragmentation time when model parameters are varied within their uncertainties (Methods).

[1]Laboratoire Lagrange, UNS-CNRS, Observatoire de la Côte d'Azur, Boulevard de l'Observatoire-CS 34229, 06304 Nice Cedex 4, France. [2]Université de Lorraine, CRPG-CNRS, 15 Rue Notre-Dame des Pauvres, BP 20, 54501 Vandoeuvre les Nancy, France. [3]Laboratoire Géoazur, UNS-CNRS, Observatoire de la Côte d'Azur, 250 rue Albert Einstein, Les Lucioles 1, Sophia-Antipolis, 06560 Valbonne, France. [4]Hopkins Extreme Materials Institute, Johns Hopkins University, Latrobe 122, 3400 North Charles Street, Baltimore, Maryland 21218, USA. [5]Institut Supérieur de l'Aéronautique et de l'Espace, 10 avenue Edouard-Belin, BP 54032, 31055 Toulouse Cedex 4, France. [6]Solar System Exploration Research Virtual Institute, Institute for the Science of Exploration Targets, Southwest Research Institute, 1050 Walnut Street, Suite 300 Boulder, Colorado 80302, USA.

fatigue crack growth, we perform laboratory experiments (Methods) on two meteorites: a carbonaceous chondrite (CM2 Murchison) and an ordinary chondrite (L/LL3.2 Sahara 97210) considered the closest available analogues of the broad asteroid spectroscopic classes C and S, respectively. Our protocol (Extended Data Fig. 1) consists of using a climatic chamber to subject these meteorites to temperature cycles that approximate the day–night temperature variations experienced on NEA surfaces. The temperature cycle period was taken to be 2.2 h, the fastest period permitted by our climatic chamber, allowing us to have a reasonable number of cycles within a month (still subjecting meteorites to temperature rates of change typical of NEA surfaces). The magnitude of the temperature excursion, ΔT, was taken to be 190 K, equal of the ΔT of C-type NEAs at ~0.7 AU from the Sun (Methods and Extended Data Fig. 2).

After subjecting the meteorites to as few as 407 temperature cycles, we use X-ray tomography to observe (Fig. 2) and measure (Methods) an increase in the length and in the width of several of the pre-existing cracks in both Murchison and Sahara 97210, thus confirming thermal

fatigue (measurements for 21 cracks for each meteorite are shown in Extended Data Figs 3 and 4). Additionally, we found particles in the sample holder of Murchison that had broken off from its surface as a result of temperature cycling (Fig. 3). Crack growth and fragmentation due to either the manipulation of the samples or the freeze–thaw of water are both ruled out (Methods and Extended Data Fig. 5).

A second question is how much time is required for pre-existing cracks to propagate enough to break the rocks and contribute to the generation of finer regolith. Our measurements show that, under laboratory conditions, pre-existing cracks are extending in length at a rate of about 0.5 mm yr^{-1} (average of Murchison and Sahara 97210). A constant crack propagation rate would suggest that a 1-cm rock on the surface of an NEA could fragment into smaller pieces in less than 20 yr, several orders of magnitude faster than comminution by micrometeoroid impacts, which requires about 2 Myr on NEAs and about 20 Myr on MBAs. However, the crack propagation rate is typically a nonlinear function of crack size, thus requiring a detailed fracture mechanics analysis to investigate whether such a rapid growth rate would be maintained to the point of fragmentation and whether thermal fragmentation can also occur for different cycle periods, for larger rocks and at lower temperatures (for example for MBAs).

To investigate this question, we develop a micromechanical model (Methods) based on well-established thermal diffusion[15], thermomechanical[16] (Extended Data Fig. 6) and fracture mechanics[17] models in order to analyse the progressive crack growth from the early stages to final fragmentation (model parameter values are in Extended Data Table 1).

First we compare the crack growth measured in our laboratory experiments with model predictions. We use thermal boundary conditions simulating those of the laboratory experiments: the entire surface of a
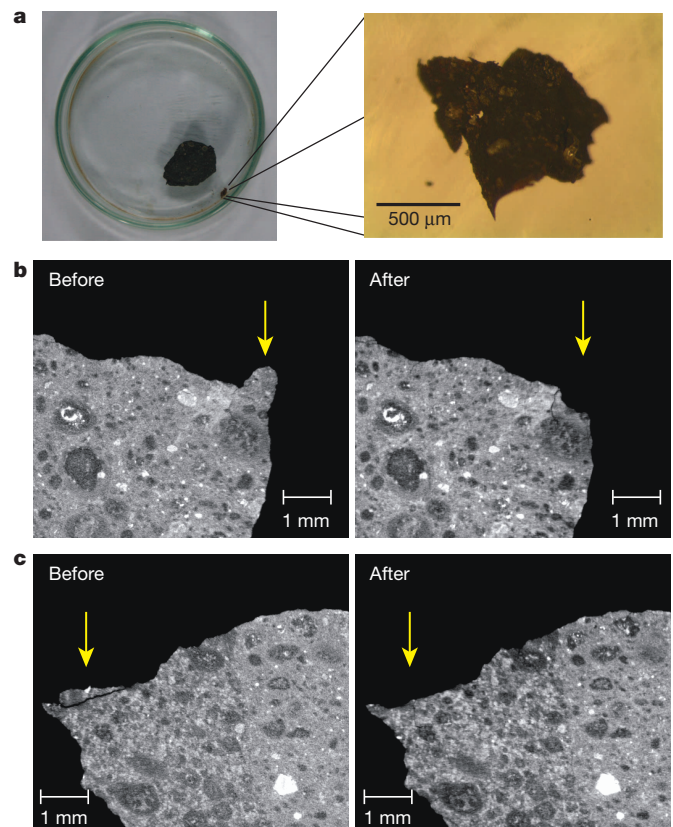
Figure 2 | Crack growth in meteorites due to laboratory temperature cycling. Arrows of the same colour mark growing cracks in the pictures before and after temperature cycling. **a**, Murchison; **b**, Sahara 97210. The cracks volume and length growth are given in the Extended Data Figs 3 and 4, respectively. **c**, The mean change of the length of the 21 measured cracks for each meteorite as a function of the number of thermal cycles. Statistical errors are 1σ.

Figure 3 | Regolith formation from Murchison in the laboratory. a, The sample of Murchison and (enlarged) one of its small fragments, containing several visible chondrules, found in the sample holder after temperature cycling. **b, c**, Tomographic slices of regions of the same sample of Murchison before and after temperature cycling. The arrows indicate fragments that broke off from Murchison.

rock 1 cm in diameter is forced to follow sinusoidal temperature oscillations with $\Delta T = 190$ K and a period of 2.2 h. Considering the idealized nature of the model, the agreement with the experimental measurements is satisfactory (the average model-to-experiment discrepancy is 20%; see Methods and Extended Data Fig. 7) and indicates that we have captured the essential aspects of thermal fatigue crack growth.

Next we use the model to provide predictions for the time (number of day–night temperature cycles) required to achieve thermal fragmentation for surface rocks with sizes between 1 and 10 cm (Extended Data Fig. 8) on C- and S-type asteroids at 1 and 2.5 AU from the Sun. Here the cycle period is taken to be 6 h, which is more appropriate than 2.2 h for most small asteroids. The day–night temperature variations, temperature gradients and mechanical stresses are calculated using boundary conditions appropriate for asteroid surfaces radiatively heated by the Sun (Methods). We use a conservative definition of rock fragmentation in our model, namely that an initial 30-μm-long crack grows to a length equal to the rock diameter. Shorter cracks can still produce fragments, either by merging with other growing cracks or by a flaking mechanism (as in the case of Fig. 3).

Our results (Fig. 1) demonstrate that, at 1 AU from the Sun, centimetre-sized rocks fragment on asteroids at least an order of magnitude faster by thermal fragmentation than by comminution by micrometeoroid impacts, the previously assumed dominant mechanism. We also find (Fig. 1) that although larger rocks require more time to fragment by micrometeoroid impact, the trend is reversed for thermal fragmentation. Therefore, thermal fragmentation of a 10-cm rock is predicted to occur orders of magnitude faster than fragmentation due to micrometeoroid impact. Although the speed of thermal fatigue is reduced at larger heliocentric distances (Extended Data Fig. 2), we still find (Fig. 1) that in the main belt, 2.5 AU from the Sun, thermal fatigue fragmentation is, in general, and within the errors of our model (Methods and Extended Data Fig. 9), more quickly than rock break-up by micrometeoroid impacts.

Because regolith formation by thermal fragmentation does not depend on asteroid size, this process occurs also on larger asteroids (for example, those varying in size from several tens to several hundreds of kilometres).

We argue that asteroid rock thermal fragmentation has observable implications: by breaking down rocks into smaller pieces and thus exposing new surface area, thermal fragmentation can provide a mechanism to make fresh regolith. Because the process of thermal fragmentation is strongly dependent on the value of ΔT, the rate of thermal fragmentation increases with decreasing perihelion distance (Extended Data Fig. 2). Observations show that the fraction of NEAs with fresh surfaces, whose reflectance spectra resemble those of ordinary chondrites ('Q-type' asteroids), is increasing with decreasing perihelion distance with a dependence[18] that mimics the curves of the NEA ΔT. From this result, a concrete, testable prediction is that more regolith should be found on NEAs that, during their chaotic dynamical evolution, are likely to have spent more time with smaller perihelion distances[19].

We also predict that small NEAs (for example those 100 m in radius) with low perihelion distances ($q < 0.3$ AU) could be eroded by thermal fragmentation and radiation pressure sweeping[20] on timescales shorter than their dynamical lifetime. For instance, at 0.3 AU the solar radiation pressure can remove grains with radii of the order of millimetres from the surface of an asteroid with a radius of 100 m (ref. 20); these grains can be produced by thermal fragmentation of centimetre-sized rocks in less than ∼200 yr. Therefore, low-perihelion NEAs loose regolith at a rate of roughly 5×10^{-5} m yr^{-1}, implying that an object with a radius of 100 m would be completely eroded in about 2 Myr. Because thermal fragmentation is faster for C-type NEAs than for S-type NEAs (Fig. 1), we predict that erosion would be faster for carbonaceous NEAs, providing an explanation for the shortage of low-albedo, carbonaceous Aten-type NEAs[10].

METHODS SUMMARY

Rock comminution by micrometeoroid impacts on asteroids. We calculate the impact probability of micrometeoroids of the zodiacal dust cloud[13] on the Moon,

on a typical NEA and on MBAs using the method of ref. 14 but without imposing any impact velocity cut-off. The ratio of the rock survival times on NEAs and MBAs to the known lunar value[12] is proportional to the respective impact probability ratios.

Thermal fatigue laboratory experiments. We perform two sets of temperature cycles (respectively 76 and 331, for a total of 407) on centimetre-sized samples of the meteorites Murchison (CM2) and Sahara 97210 (L/LL3.2). Before and after the first set of cycles, and after the second set, meteorites are imaged using a computed tomographic scanner. The scanned volumes are aligned and the increases in volume and length of several (21) cracks are measured.

Thermomechanical and fatigue crack growth model. Asteroid rock (Extended Data Fig. 8) temperatures are calculated for heliocentric distances between 0.3 and 2.5 AU. We use a well-established one-dimensional thermal model[15]. Model parameters are as follows[21]: rotational period, 6 h; thermal inertia, $\Gamma = 640$ J m^{-2} s$^{-0.5}$ K^{-1} (carbonaceous chondrite) or $\Gamma = 1{,}800$ J m^{-2} s$^{-0.5}$ K^{-1} (ordinary chondrite); bolometric albedo, $A = 0.02$ (carbonaceous) or $A = 0.1$ (ordinary). The cyclic spatial and temporal temperature fields are used to calculate the strains and stresses in the rocks and the stress intensity factors at the tip of surface cracks that, following Paris's law[17], grow downwards through the rocks in the direction perpendicular to the asteroid surface. The time needed to propagate a crack from 30 μm to a length equal to the rock diameter is then compared with the rock survival time against micrometeoroid impact (Fig. 1).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Veverka, J. *et al.* The landing of the NEAR-Shoemaker spacecraft on asteroid 433 Eros. *Nature* **413,** 390–393 (2001).
2. Yano, H. *et al.* Touchdown of the Hayabusa spacecraft at the Muses Sea on Itokawa. *Science* **312,** 1350–1353 (2006).
3. Gundlach, B. & Blum, J. A new method to determine the grain size of planetary regolith. *Icarus* **223,** 479–492 (2013).
4. Housen, K. R., Wilkening, L. L., Chapman, C. R. & Greenberg, R. Asteroidal regoliths. *Icarus* **39,** 317–351 (1979).
5. Hörz, F. & Cintala, M. Impact experiments related to the evolution of planetary regoliths. *Meteorit. Planet. Sci.* **32,** 179–209 (1997).
6. Housen, K. R. & Holsapple, K. A. Ejecta from impact craters. *Icarus* **211,** 856–875 (2011).
7. McFadden, L., Eppes, M., Gillespie, A. & Hallet, B. Physical weathering in arid landscapes due to diurnal variation in the direction of solar heating. *Geol. Soc. Am. Bull.* **117,** 161–173 (2005).
8. Luque, A., Ruiz-Agudo, E., Cultrone, G., Sebastián, E. & Siegesmund, S. Direct observation of microcrack development in marble caused by thermal weathering. *Environ. Earth Sci.* **62,** 1375–1386 (2011).
9. Viles, H. *et al.* Simulating weathering of basalt on mars and earth by thermal cycling. *Geophys. Res. Lett.* **37,** L18201 (2010).
10. Mainzer, A. *et al.* Characterizing subpopulations within the near-earth objects with NEOWISE: preliminary results. *Astrophys. J.* **752,** 110–126 (2012).
11. Michel, P. & Richardson, D. C. Collision and gravitational reaccumulation: possible formation mechanism of the asteroid Itokawa. *Astron. Astrophys.* **554,** L1 (2013).
12. Hoerz, F., Schneider, E., Gault, D. E., Hartung, J. B. & Brownlee, D. E. Catastrophic rupture of lunar rocks: a Monte Carlo simulation. *Moon* **13,** 235–258 (1975).
13. Nesvorný, D. *et al.* Cometary origin of the zodiacal cloud and carbonaceous micrometeorites. implications for hot debris disks. *Astrophys. J.* **713,** 816–836 (2010).
14. Briani, G., Morbidelli, A., Gounelle, M. & Nesvorný, D. Evidence for an asteroid-comet continuum from simulations of carbonaceous microxenolith dynamical evolution. *Meteorit. Planet. Sci.* **46,** 1863–1877 (2011).
15. Spencer, J. R., Lebofsky, L. A. & Sykes, M. V. Systematic biases in radiometric diameter determinations. *Icarus* **78,** 337–354 (1989).
16. Čapek, D. & Vokrouhlický, D. Thermal stresses in small meteoroids. *Astron. Astrophys.* **519,** A75 (2010).
17. Janssen, M., Zuidema, J. & Wanhill, R. *Fracture Mechanics* 2nd edn, 207–213 (Spon, 2004).
18. Marchi, S., Magrin, S., Nesvorný, D., Paolicchi, P. & Lazzarin, M. A spectral slope versus perihelion distance correlation for planet-crossing asteroids. *Mon. Not. R. Astron. Soc. Lett.* **368,** L39–L42 (2006).
19. Marchi, S., Delbo, M., Morbidelli, A., Paolicchi, P. & Lazzarin, M. Heating of near-Earth objects and meteoroids due to close approaches to the sun. *Mon. Not. R. Astron. Soc.* **400,** 147–153 (2009).
20. Jewitt, D. The active asteroids. *Astron. J.* **143,** 66–80 (2012).
21. Opeil, C. P., Consolmagno, G. J. & Britt, D. T. The thermal conductivity of meteorites: new measurements and analysis. *Icarus* **208,** 449–454 (2010).

LETTER

A quantum gate between a flying optical photon and a single trapped atom

Andreas Reiserer[1], Norbert Kalb[1], Gerhard Rempe[1] & Stephan Ritter[1]

The steady increase in control over individual quantum systems supports the promotion of a quantum technology that could provide functionalities beyond those of any classical device. Two particularly promising applications have been explored during the past decade: photon-based quantum communication, which guarantees unbreakable encryption[1] but which still has to be scaled to high rates over large distances, and quantum computation, which will fundamentally enhance computability[2] if it can be scaled to a large number of quantum bits (qubits). It was realized early on that a hybrid system of light qubits and matter qubits[3] could solve the scalability problem of each field—that of communication by use of quantum repeaters[4], and that of computation by use of an optical interconnect between smaller quantum processors[5,6]. To this end, the development of a robust two-qubit gate that allows the linking of distant computational nodes is "a pressing challenge"[6]. Here we demonstrate such a quantum gate between the spin state of a single trapped atom and the polarization state of an optical photon contained in a faint laser pulse. The gate mechanism presented[7,8] is deterministic and robust, and is expected to be applicable to almost any matter qubit. It is based on reflection of the photonic qubit from a cavity that provides strong light–matter coupling. To demonstrate its versatility, we use the quantum gate to create atom–photon, atom–photon–photon and photon–photon entangled states from separable input states. We expect our experiment to enable various applications, including the generation of atomic[9] and photonic[10] cluster states and Schrödinger-cat states[11], deterministic photonic Bell-state measurements[12], scalable quantum computation[7] and quantum communication using a redundant quantum parity code[13].

Since their infancy, the fields of quantum communication and quantum computation have been largely independent. For communication[1], optical photons are used because they allow the transmission of quantum states, such as time-bin or polarization qubits, over large distances using existing telecommunication fibre technology. Quantum computation[2], on the other hand, is typically based on single spins, either in vacuum or in specific solid-state host materials. In addition to the long coherence times that these spins can exhibit, they provide deterministic interaction mechanisms that facilitate local two-qubit quantum gates. Scalability would be offered by combining the specific advantages of both information carriers—namely, spins and photons[5,6]. To implement the required interaction between the different types of qubits, a deterministic quantum gate between a photon and an atom has been proposed[7]. Here we demonstrate this quantum gate and its potential for quantum information processing with atoms and photons.

The mechanism[8] we use is based on cavity quantum electrodynamics. When a photon interacts with a cavity containing a single, resonant emitter, it experiences a phase shift[14,15] which depends on the coupling strength. In our experiment, the emitter is a single ^{87}Rb atom, which is trapped at the centre of an overcoupled cavity. Full control over the position and motion of the atom[16] puts the system into the strong-coupling regime (measured coupling constant $g = 2\pi \times 6.7\,\mathrm{MHz}$, atomic dipole decay rate $\gamma = 2\pi \times 3.0\,\mathrm{MHz}$, cavity field decay rate $\kappa = 2\pi \times 2.5\,\mathrm{MHz}$). In this regime, the conditional phase shift induced on

a reflected light field[8] is π, which is the prerequisite for the quantum gate presented in this work.

In contrast to the original proposal[7], our implementation does not require interferometric stability, as the a.c. Stark shift of a linearly polarized dipole trap is used to split the Zeeman states of the excited atomic state manifold (see Methods and the level scheme in Fig. 1a). Thus, the coupling is only strong when the atom (a) is in state $|\uparrow^a\rangle$ and photons (p) of right-circular polarization $|\uparrow^p\rangle$ are reflected (green arrow and sphere in Fig. 1a). For all other qubit combinations (red arrows and sphere), the coupling is negligible because any atomic transition is detuned (see Methods). Therefore, the reflection of a photon results in a conditional phase shift of π, that is, a sign change, between the atomic and the photonic qubit:

$$|\uparrow^a\uparrow^p\rangle \rightarrow |\uparrow^a\uparrow^p\rangle$$
$$|\downarrow^a\uparrow^p\rangle \rightarrow -|\downarrow^a\uparrow^p\rangle$$
$$|\uparrow^a\downarrow^p\rangle \rightarrow -|\uparrow^a\downarrow^p\rangle$$
$$|\downarrow^a\downarrow^p\rangle \rightarrow -|\downarrow^a\downarrow^p\rangle$$

This conditional phase shift allows the construction of a universal quantum gate that can be transformed into any two-qubit gate using rotations of the individual qubits, which are implemented with wave plates for the photon and with Raman transitions for the atom. With respect to the photonic basis states $|\uparrow^p_x\rangle \equiv \frac{1}{\sqrt{2}}(|\uparrow^p\rangle + |\downarrow^p\rangle)$ and $|\downarrow^p_x\rangle \equiv \frac{1}{\sqrt{2}}(|\uparrow^p\rangle - |\downarrow^p\rangle)$, the conditional phase shift represents an atom–photon controlled-NOT (CNOT) gate.

The action of the quantum CNOT gate is a flip of the photonic target qubit, controlled by the quantum state of the atom, similar to its classical analogue. A first step to characterize the gate is therefore to measure a classical truth table. To this end, the atomic state is prepared by optical pumping either into the uncoupled $F = 1$ states, corresponding to $|\downarrow^a\rangle$, or into the coupled $|\uparrow^a\rangle$ state (see Methods). Subsequently, faint laser pulses (average photon number $\bar{n} = 0.3$) in $|\downarrow^p_x\rangle$ or $|\uparrow^p_x\rangle$ are reflected from the cavity and measured with single-photon counting modules in a polarization-resolving set-up. To ensure spectral mode matching[7], we use a Gaussian photon wavepacket with a full-width at half-maximum (FWHM) of 0.7 μs, corresponding to a FWHM bandwidth of 0.6 MHz, which is almost an order of magnitude smaller than the cavity FWHM linewidth of 5 MHz. After the reflection process, the atomic state is measured within 3 μs using cavity-enhanced hyperfine-state detection (see Methods, Extended Data Fig. 1 and ref. 8). The results are shown in Fig. 1b (see also Extended Data Table 1a), where the bars represent the normalized probabilities of detecting a certain output state for each of the orthogonal input states.

The control and target qubits are expected to be unchanged when the control qubit is in the state $|\downarrow^a\rangle$, which is accomplished with a probability of 99%. This number is limited by imperfections in the
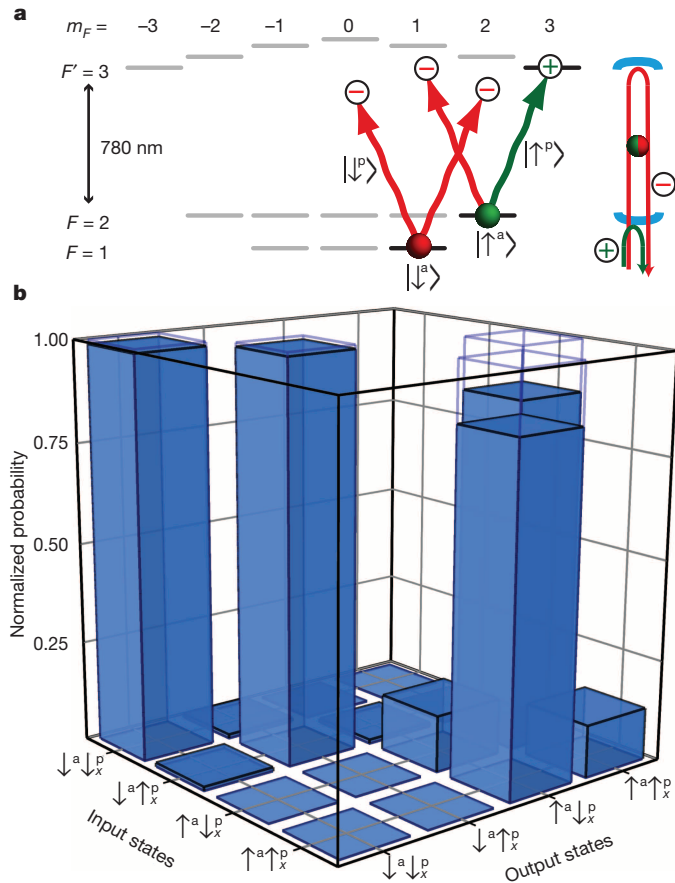
Figure 1 | Atom–photon quantum gate. a, Atomic level scheme. The photonic qubit basis states are left-circular ($|\downarrow^{\mathrm{P}}\rangle$) and right-circular ($|\uparrow^{\mathrm{P}}\rangle$) polarization. The atomic qubit is encoded in the $|F, m_F\rangle$ states $|\downarrow^{\mathrm{a}}\rangle \equiv |1,1\rangle$ and $|\uparrow^{\mathrm{a}}\rangle \equiv |2,2\rangle$. Here, F denotes the hyperfine state and m_F its projection onto an external magnetic field. The cavity (blue semicircles) is resonant with the a.c. Stark-shifted $|2,2\rangle \leftrightarrow |3,3\rangle$ transition. On reflection of a photon, the combined atom–photon state $|\uparrow^{\mathrm{a}}\uparrow^{\mathrm{P}}\rangle$ (green, $\oplus$) acquires a phase shift of π with respect to all other states (red, $\ominus$). **b**, Measured truth table, showing the normalized probability of obtaining a certain output state for a complete orthogonal set of input states. Open blue bars indicate an ideal CNOT gate.

detection of the photon polarization and the atomic hyperfine state. When the control qubit is in $|\uparrow^{\mathrm{a}}\rangle$, the expected flip of the photonic target qubit is observed with a probability of 86%. The statistical errors in the depicted data are negligible. However, we observe ambient-temperature-related drifts of about 2% on a timescale of several hours. The flip probability is predominantly limited by three effects: first, by optical mode matching, because the transverse overlap between the free-space mode of the photon and the cavity mode is 92(3)%. Second, by the quality of preparation of the state $|\uparrow^{\mathrm{a}}\rangle$, which is successful with 96(1)% probability. Last, by the stability of the difference between the cavity resonance and the frequency of the impinging laser pulse, which is about 300 kHz. None of these imperfections has a fundamental limit.

The decisive feature that distinguishes a quantum gate from a classical one is the generation of entangled states from separable input states. To characterize this property, faint laser pulses ($\bar{n} = 0.07$, FWHM 0.7 µs) are reflected from the set-up and the evaluation is post-selected on those cases where a single photon has subsequently been detected. The input state is $|\downarrow^{\mathrm{a}}_x\downarrow^{\mathrm{P}}_x\rangle$, such that the gate generates the maximally entangled $|\Phi^+_{\mathrm{ap}}\rangle$ state:

$$|\downarrow^{\mathrm{a}}_x\downarrow^{\mathrm{P}}_x\rangle \rightarrow |\Phi^+_{\mathrm{ap}}\rangle = \frac{1}{\sqrt{2}}\left(|\downarrow^{\mathrm{a}}\downarrow^{\mathrm{P}}\rangle + |\uparrow^{\mathrm{a}}\uparrow^{\mathrm{P}}\rangle\right)$$

Both the atomic and the photonic qubits are measured in three orthogonal bases. This allows us to reconstruct the density matrix ρ_{ap} of

the combined atom–photon state using quantum-state tomography and a maximum-likelihood estimation[17]. The result is shown in Fig. 2 (see also Extended Data Table 1b). In accordance with the truth table measurement above, the density matrix is slightly asymmetric. Whereas the value of $|\downarrow^{\mathrm{a}}_x\downarrow^{\mathrm{P}}_x\rangle\langle\downarrow^{\mathrm{a}}_x\downarrow^{\mathrm{P}}_x|$ (left corner) is close to the ideal 0.5, the elements in the other corners are smaller. The fidelity with the expected $|\Phi^+_{\mathrm{ap}}\rangle$ state is $F_{\Phi^+_{\mathrm{ap}}} = \langle\Phi^+_{\mathrm{ap}}|\rho_{\mathrm{ap}}|\Phi^+_{\mathrm{ap}}\rangle = 80.7(0.5)\%$, where the standard error has been determined with the Monte Carlo technique[17]. In the depicted measurement, the fidelity with a slightly rotated, maximally entangled state of the form $\frac{1}{\sqrt{2}}\left(|\uparrow^{\mathrm{a}}\uparrow^{\mathrm{P}}_x\rangle + e^{-i\varphi}|\downarrow^{\mathrm{a}}\downarrow^{\mathrm{P}}_x\rangle\right)$ can be higher, probably due to a small frequency offset between the cavity and the photon. We find a maximum value of 83.0% for $\varphi = 0.11\pi$.

The major experimental imperfections that reduce the fidelity are as follows (see Methods for details): first, the above-mentioned spatial mode mismatch between cavity and impinging photon (reduction 8(3)%); second, the quality of our atomic state preparation, rotation and readout (see Extended Data Fig. 2; reduction 5(1)%); third, imperfections in the photonic polarization measurement and detector dark counts (reduction 3%); and last, the small probability of having more than one photon in the impinging laser pulses (reduction 2%). Again, none of these limitations is fundamental.

In principle, the gate mechanism presented in this work is deterministic. In our experimental implementation, the photon is not back-reflected[8] from the coupled system $|\uparrow^{\mathrm{a}}\uparrow^{\mathrm{P}}\rangle$ with a probability of 34(2)% (due to the finite cooperativity $C = \frac{g^2}{2\kappa\gamma} = 3$) and in the uncoupled cases with a probability of 30(2)% (due to the non-zero transmission of the highly reflecting cavity mirror and the mirror scattering and absorption losses, see ref. 8 and Methods). The small difference in reflectivity also contributes slightly (<1%) to the observed reduction in fidelity[9]. The achieved loss level nevertheless allows for scalable quantum computation[18] and deterministic quantum state transfer[19]. One would still observe non-classical correlations without post-selection if a perfect single-photon source and a perfect detector were used
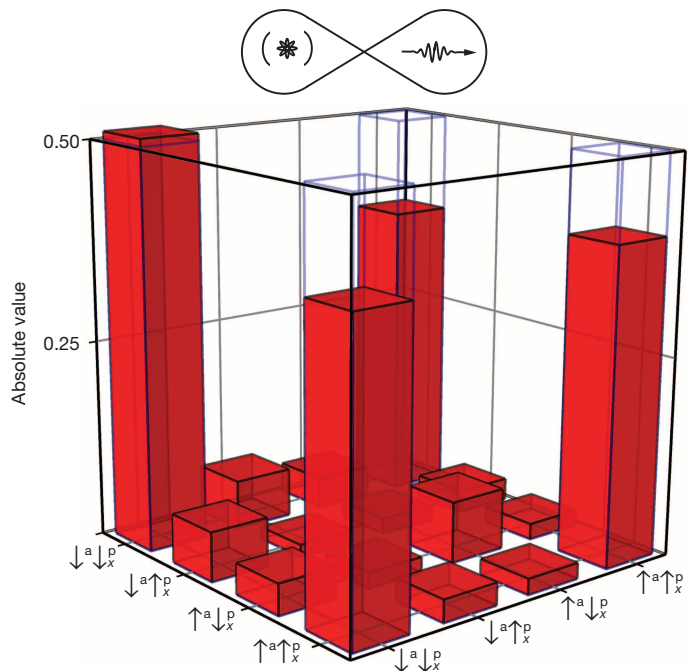

Figure 2 | Entangled atom–photon state generated via the gate operation. The bars show the absolute value of the density-matrix elements. The fidelity with the maximally entangled $|\Phi^+_{\mathrm{ap}}\rangle$ Bell state (open blue bars) is 80.7(0.5)%. The diagram at the top of the figure symbolizes atom–photon entanglement.
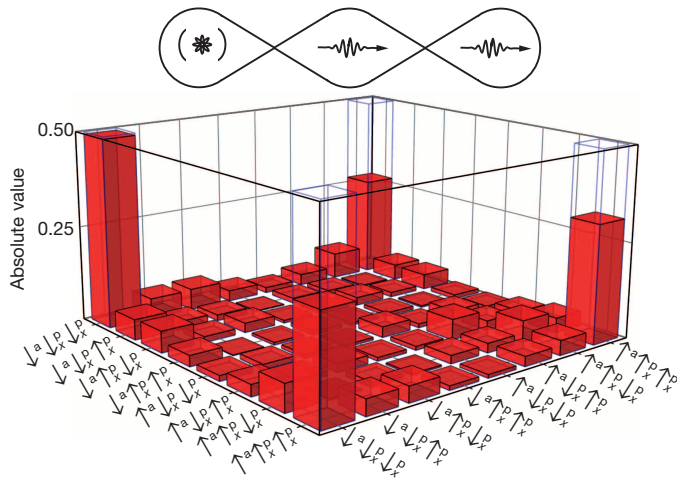
Figure 3 | Entangled state between one atom and two photons. This state is generated by reflecting two faint laser pulses from the cavity. The bars show the absolute value of the reconstructed density matrix elements. The fidelity with the maximally entangled state $|\mathrm{GHZ}\rangle$ is 61(2)%. The matrix elements of $|\mathrm{GHZ}\rangle$ are depicted as open blue bars. The diagram at the top of the figure symbolizes entanglement between an atom and two photons.

to characterize our device. We also expect that it will be possible to dramatically improve the achieved gate efficiency in next-generation cavities with increased atom–cavity coupling strength[20–23] and reduced losses.

The demonstrated quantum gate also allows the generation of entangled cluster states that consist of a trapped atom and several flying photons; this is complementary to experiments with flying atoms and trapped microwave photons[24]. To this end, the gate is applied to the photons contained in two sequentially impinging laser pulses (temporal separation 3 µs). Post-selecting events where one photon was detected in each of the input pulses, a maximally entangled Greenberger–Horne–Zeilinger (GHZ) state is expected:

$$\left|\downarrow_x^a \downarrow_x^P \downarrow_x^P\right\rangle \to |\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}\left(\left|\uparrow^a \uparrow_x^P \uparrow_x^P\right\rangle - \left|\downarrow^a \downarrow_x^P \downarrow_x^P\right\rangle\right)$$

The density matrix of the generated quantum state, again reconstructed using quantum state tomography and a maximum-likelihood estimation, is shown in Fig. 3 (see also Extended Data Table 1c). The fidelity with the ideal state $|\mathrm{GHZ}\rangle$ (open blue bars in Fig. 3) is 61(2)%, proving genuine three-particle (atom–photon–photon) entanglement. The reasons for a non-unity fidelity are analogous to those for the case of two particles. Again, we experimentally find a higher fidelity of 67% with the slightly rotated GHZ state $\frac{1}{\sqrt{2}}\left(\left|\uparrow^a \uparrow_x^P \uparrow_x^P\right\rangle - e^{-i\varphi}\left|\downarrow^a \downarrow_x^P \downarrow_x^P\right\rangle\right)$, with $\varphi = 0.21\pi$.

Finally, we investigate whether the presented gate mechanism can mediate a photon–photon interaction for optical quantum computing[7]. We employ a quantum eraser protocol[10] which should allow us to create a maximally entangled state out of two separable input photons. To this end, the state $|\mathrm{GHZ}\rangle$ is generated as described above and a $\pi/2$ rotation is applied to the atom, which transforms the state to:

$$\frac{1}{\sqrt{2}}\left[\left|\uparrow^a\right\rangle\left(\left|\uparrow_x^P \uparrow_x^P\right\rangle - \left|\downarrow_x^P \downarrow_x^P\right\rangle\right) - \left|\downarrow^a\right\rangle\left(\left|\uparrow_x^P \uparrow_x^P\right\rangle + \left|\downarrow_x^P \downarrow_x^P\right\rangle\right)\right]$$

Subsequent measurement of the atomic state disentangles the atom, which results in a maximally entangled two-photon state: if the atom is found in $|\downarrow^a\rangle$ ($|\uparrow^a\rangle$), the resulting state is $\left|\Phi_{PP}^+\right\rangle$ $\left(\left|\Phi_{PP}^-\right\rangle\right)$, respectively. In

the experiment, the two-photon density matrices are again reconstructed with the maximum-likelihood technique (see Fig. 4 and Extended Data Table 1d and e). This gives a fidelity with the expected Bell states of 67(2)% (64(2)%) for the $\left|\Phi_{PP}^+\right\rangle$ $\left(\left|\Phi_{PP}^-\right\rangle\right)$ state. The values achieved prove photon–photon entanglement. Their small difference can be explained by the fact that a detection of the atom in $F = 1$ selects only those events where it has initially been prepared in the correct state $|\uparrow^a\rangle$, rather than in another state of the $F = 2$ hyperfine manifold. Again, we find a higher fidelity of maximally 76% with a rotated $\left|\Phi_{PP}^+\right\rangle$ state with $\varphi = 0.25\pi$.

The above measurements demonstrate the versatility of the presented gate mechanism and its ability to mediate a photon–photon interaction. To this end, intermediate storage of the two photons during the time required to rotate and read out the atomic state (about 3 µs) is required, which can be implemented with an optical fibre less than one kilometre in length. Conditioned on the state of the atom, the polarization of the photons then has to be rotated using, for example, an electro-optical modulator. As an alternative to the eraser scheme used in this work, the first photon could be reflected from the cavity a second time[7].

In addition to the applications mentioned above[5–13], the gate mechanism we present here opens up perspectives for numerous quantum optics experiments. First, it could be applied to perform a quantum-non-demolition measurement of the polarization of a single reflected photon by measuring the state of the atom—conversely, it could be used to measure the atomic state without energy exchange[21] by measuring the polarization of a reflected photon. Second, a quantum gate between several atoms in the same or even in remote cavities[25,26] could be directly implemented, which would facilitate universal quantum computation in a decoherence-free subspace[27]. Last, the use of our gate mechanism in the proposed deterministic optical Bell-state measurement[12] would markedly increase the efficiency of teleportation between remote atoms[28,29], and would therefore improve the prospects for implementation of a quantum repeater[4] and a global-scale quantum network[3,30].
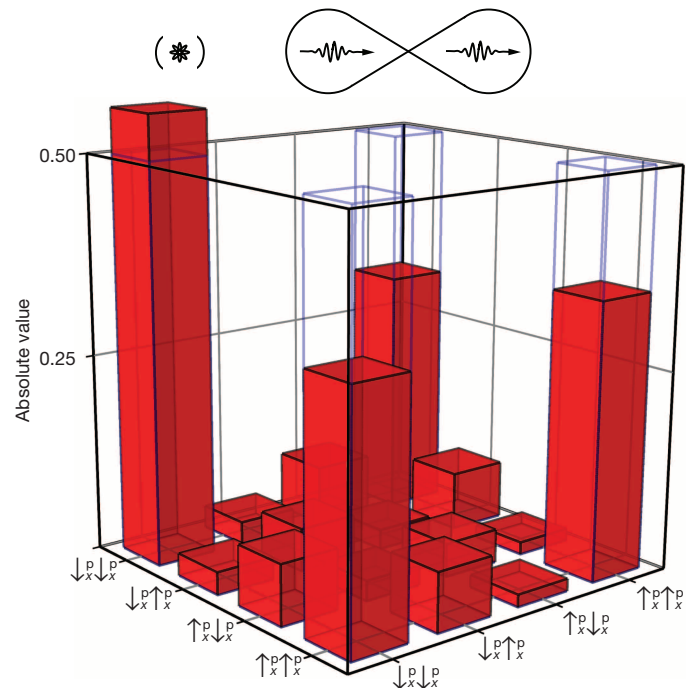

Figure 4 | Entangled photon–photon state generated via consecutive interaction with the atom. The bars show the absolute value of the density matrix elements. The fidelity with the maximally entangled $\left|\Phi_{PP}^+\right\rangle$ Bell state (open blue bars) is 67(2)%. The diagram at the top of the figure symbolizes photon–photon entanglement.

METHODS SUMMARY

In the experimental set-up, single ^{87}Rb atoms are loaded from a magneto-optical trap into a three-dimensional optical lattice inside a Fabry–Perot cavity. The cavity is overcoupled, that is, the coupling mirror has a transmission (95 p.p.m.) which is large compared to the transmission of the other mirror in the cavity and the scattering and absorption losses (8 p.p.m. total). The geometry of the trap and the cooling mechanisms used are described in detail in ref. 16. The lattice consists of three retro-reflected laser beams, one red-detuned (1,064 nm) and two blue-detuned (770 nm) from the atomic transitions at 780 nm (D_2 line) and 795 nm (D_1 line). The use of high intensities leads to trap frequencies of several hundred kHz, which facilitates fast cooling to low temperatures using intra-cavity Sisyphus cooling. In each experimental cycle, a cooling interval of 0.8 ms is applied, which allows for atom trapping times of many seconds. In contrast to ref. 16, ground-state cooling is not applied in this work.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74,** 145–195 (2002).
2. Ladd, T. D. *et al.* Quantum computers. *Nature* **464,** 45–53 (2010).
3. Duan, L.-M. & Monroe, C. Colloquium: Quantum networks with trapped ions. *Rev. Mod. Phys.* **82,** 1209–1224 (2010).
4. Briegel, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81,** 5932–5935 (1998).
5. Monroe, C. & Kim, J. Scaling the ion trap quantum processor. *Science* **339,** 1164–1169 (2013).
6. Awschalom, D. D., Bassett, L. C., Dzurak, A. S., Hu, E. L. & Petta, J. R. Quantum spintronics: engineering and manipulating atom-like spins in semiconductors. *Science* **339,** 1174–1179 (2013).
7. Duan, L.-M. & Kimble, H. J. Scalable photonic quantum computation through cavity-assisted interactions. *Phys. Rev. Lett.* **92,** 127902 (2004).
8. Reiserer, A., Ritter, S. & Rempe, G. Nondestructive detection of an optical photon. *Science* **342,** 1349–1351 (2013).
9. Cho, J. & Lee, H.-W. Generation of atomic cluster states through the cavity input-output process. *Phys. Rev. Lett.* **95,** 160501 (2005).
10. Hu, C. Y., Munro, W. J. & Rarity, J. G. Deterministic photon entangler using a charged quantum dot inside a microcavity. *Phys. Rev. B* **78,** 125318 (2008).
11. Wang, B. & Duan, L.-M. Engineering superpositions of coherent states in coherent optical pulses through cavity-assisted interaction. *Phys. Rev. A* **72,** 022320 (2005).
12. Bonato, C. *et al.* CNOT and Bell-state analysis in the weak-coupling cavity QED regime. *Phys. Rev. Lett.* **104,** 160503 (2010).
13. Munro, W. J., Stephens, A. M., Devitt, S. J., Harrison, K. A. & Nemoto, K. Quantum communication without the necessity of quantum memories. *Nature Photon.* **6,** 777–781 (2012).
14. Turchette, Q. A., Hood, C. J., Lange, W., Mabuchi, H. & Kimble, H. J. Measurement of conditional phase shifts for quantum logic. *Phys. Rev. Lett.* **75,** 4710–4713 (1995).
15. Fushman, I. *et al.* Controlled phase shifts with a single quantum dot. *Science* **320,** 769–772 (2008).
16. Reiserer, A., Nölleke, C., Ritter, S. & Rempe, G. Ground-state cooling of a single atom at the center of an optical cavity. *Phys. Rev. Lett.* **110,** 223003 (2013).
17. Paris, M. & Řeháček, J. *Quantum State Estimation* (Springer, 2004).
18. Duan, L.-M. & Raussendorf, R. Efficient quantum computation with probabilistic quantum gates. *Phys. Rev. Lett.* **95,** 080503 (2005).
19. van Enk, S. J., Cirac, J. I. & Zoller, P. Photonic channels for quantum communication. *Science* **279,** 205–208 (1998).
20. Dayan, B. *et al.* A photon turnstile dynamically regulated by one atom. *Science* **319,** 1062–1065 (2008).
21. Volz, J., Gehr, R., Dubois, G., Estéve, J. & Reichel, J. Measurement of the internal state of a single atom without energy exchange. *Nature* **475,** 210–213 (2011).
22. Thompson, J. D. *et al.* Coupling a single trapped atom to a nanoscale optical cavity. *Science* **340,** 1202–1205 (2013).
23. O'Shea, D., Junge, C., Volz, J. & Rauschenbeutel, A. Fiber-optical switch controlled by a single atom. *Phys. Rev. Lett.* **111,** 193601 (2013).
24. Rauschenbeutel, A. *et al.* Step-by-step engineered multiparticle entanglement. *Science* **288,** 2024–2028 (2000).
25. Xiao, Y.-F. *et al.* Realizing quantum controlled phase flip through cavity QED. *Phys. Rev. A* **70,** 042314 (2004).
26. Duan, L.-M., Wang, B. & Kimble, H. J. Robust quantum gates on neutral atoms with cavity-assisted photon scattering. *Phys. Rev. A* **72,** 032333 (2005).
27. Xue, P. & Xiao, Y.-F. Universal quantum computation in decoherence-free subspace with neutral atoms. *Phys. Rev. Lett.* **97,** 140501 (2006).
28. Olmschenk, S. *et al.* Quantum teleportation between distant matter qubits. *Science* **323,** 486–489 (2009).
29. Nölleke, C. *et al.* Efficient teleportation between remote single-atom quantum memories. *Phys. Rev. Lett.* **110,** 140403 (2013).
30. Ritter, S. *et al.* An elementary quantum network of single atoms in optical cavities. *Nature* **484,** 195–200 (2012).

LETTER

Nanophotonic quantum phase switch with a single atom

T. G. Tiecke[1,2]*, J. D. Thompson[1]*, N. P. de Leon[1,3], L. R. Liu[1], V. Vuletić[2] & M. D. Lukin[1]

By analogy to transistors in classical electronic circuits, quantum optical switches are important elements of quantum circuits and quantum networks[1–3]. Operated at the fundamental limit where a single quantum of light or matter controls another field or material system[4], such a switch may enable applications such as long-distance quantum communication[5], distributed quantum information processing[2] and metrology[6], and the exploration of novel quantum states of matter[7]. Here, by strongly coupling a photon to a single atom trapped in the near field of a nanoscale photonic crystal cavity, we realize a system in which a single atom switches the phase of a photon and a single photon modifies the atom's phase. We experimentally demonstrate an atom-induced optical phase shift[8] that is nonlinear at the two-photon level[9], a photon number router that separates individual photons and photon pairs into different output modes[10], and a single-photon switch in which a single 'gate' photon controls the propagation of a subsequent probe field[11,12]. These techniques pave the way to integrated quantum nanophotonic networks involving multiple atomic nodes connected by guided light.

A quantum optical switch[11,13–16] is challenging to implement because the interaction between individual photons and atoms is generally very weak. Cavity quantum electrodynamics (cavity QED), in which a photon is confined to a small spatial region and made to interact strongly with an atom, is a promising approach to overcoming this challenge[4]. Over the past two decades, cavity QED has enabled advances in the control of microwave[17–19] and optical[13,20–23] fields. Although integrated circuits with strong coupling of microwave photons to superconducting quantum bits (qubits) are being developed at the moment[24], a scalable path to integrated quantum circuits involving coherent qubits coupled by means of optical photons has yet to emerge.

Our experimental approach (Fig. 1a) makes use of a single atom trapped in the near field of a nanoscale photonic crystal cavity that is attached to an optical fibre taper[25]. The tight confinement of the optical mode to a volume $V \approx 0.4\lambda^3$, below the scale of the optical wavelength, λ, results in strong atom–photon interactions for an atom sufficiently close to the surface of the cavity. The atom is trapped about 200 nm from the surface in an optical lattice formed by the interference of an optical tweezer and its reflection from the side of the cavity (Methods Summary, Supplementary Information and Fig. 1a, b). Compared with transient coupling of unconfined atoms[13,22], trapping an atom allows for experiments exploiting long atomic coherence times, and enables scaling to quantum circuits with multiple atoms.
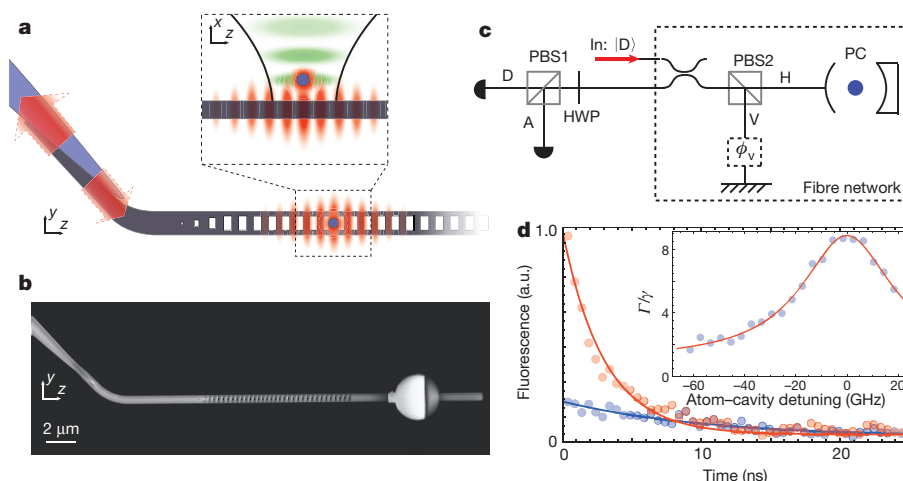

Figure 1 | Strong coupling of a trapped atom to a photonic crystal cavity. a, A single ^{87}Rb atom (blue circle) is trapped in the evanescent field (red) of a photonic crystal (grey). The photonic crystal is attached to a tapered optical fibre (blue), which provides mechanical support and an optical interface to the cavity. The tapered fibre–waveguide interface provides an adiabatic coupling of the fibre mode to the waveguide mode. The inset shows the one-dimensional trapping lattice (green), formed by the interference of a set of optical tweezers and its reflection from the photonic crystal. **b**, Scanning electron microscope (SEM) image of a single-sided photonic crystal. The pad on the right-hand side is used to tune the cavity resonance thermally by laser heating. **c**, The photonic crystal (PC) is integrated in a fibre-based polarization interferometer. A polarizing beam splitter (PBS2) splits the D-polarized input field into an H-polarized arm containing the photonic crystal and a V-polarized arm with adjustable phase ϕ_V. Using a polarizing beam splitter (PBS1) and a half-wave plate (HWP), the outgoing D and A polarizations are detected independently. **d**, Excited-state lifetime at an atom–cavity detuning of 0 (red) and −41 GHz (blue). The excited-state lifetime is shortened to $\tau = \Gamma^{-1} = 3.0(1)$ ns from the free-space value of $\gamma^{-1} = 26$ ns, yielding a cooperativity of $\eta = 7.7 \pm 0.4$. The difference in the fluorescence signal at $t = 0$ for the two detunings is consistent with the change in cavity detuning. The inset shows the enhancement of the atomic decay rate versus atom–cavity detuning. a.u., arbitrary units.

[1]Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. [2]Department of Physics and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA.
*These authors contributed equally to this work.

We use a one-sided optical cavity with a single port for both input and output[8]. In the absence of intracavity loss, photons incident on the cavity are always reflected. However, a single, strongly coupled atom changes the phase of the reflected photons by π relative to an empty cavity. More specifically, in the limit of low incident intensity, the amplitude reflection coefficient of the atom–cavity system is given by[26]

$$r_c(\eta) = \frac{(\eta - 1)\gamma + 2i\delta}{(\eta + 1)\gamma - 2i\delta} \qquad (1)$$

where $\eta = (2g)^2/\kappa\gamma$ is the cooperativity, $2g$ is the single-photon Rabi frequency, δ is the atom–photon detuning and the cavity is taken to be resonant with the driving laser. In our apparatus, the cavity intensity and atomic population decay rates are given by $\kappa = 2\pi \times 25$ GHz and $\gamma = 2\pi \times 6$ MHz, respectively. The reflection coefficient in equation (1) changes sign depending on the presence ($\eta > 1$) or absence ($\eta = 0$) of a strongly coupled atom. If the atom is prepared in a superposition of internal states, one of which does not couple to the cavity mode (for example another hyperfine atomic sublevel), the phase of the atomic superposition is switched by π on reflection of a single photon. By also adding an auxiliary photon mode that does not enter the cavity (for example an orthogonal polarization), this operation can be used to realize the Duan–Kimble scheme for a controlled-phase gate between an atomic and a photonic quantum bit[8]. The property of the atom–cavity system that a single photon and a single atom can switch each other's phase by π is the key feature of this work.

We quantify the single-atom cooperativity, η, by measuring the lifetime, τ, of the atomic excited state when it is coupled to the cavity. We excite the atom with a short (3-ns) pulse of light co-propagating with the optical trap and resonant with the $|5S_{1/2}, F = 2\rangle \rightarrow |5P_{3/2}, F' = 3\rangle$ transition (near 780 nm). The atomic fluorescence is collected through the cavity to determine the reduced excited-state lifetime, $\tau = \Gamma^{-1}$ (Fig. 1d), which yields the cooperativity, $\eta = (\Gamma - \gamma)/\gamma$. Fitting a single exponential decay gives $\tau = 3.0 \pm 0.1$ ns, corresponding to $\eta = 7.7 \pm 0.3$ and a single-photon Rabi frequency of $2g = 2\pi \times (1.09 \pm 0.03)$ GHz.

To probe the optical phase shift resulting from the atom–photon interaction, we integrate the cavity into a fibre-based polarization interferometer, which converts phase shifts into polarization rotations (Fig. 1c). The horizontally (H) polarized arm of the interferometer contains the cavity, and the vertically (V) polarized arm is used as a phase reference. For an input photon state $|\psi_{in}\rangle$ in the polarization basis $\{|H\rangle, |V\rangle\}$, the state exiting the interferometer is given by $R|\psi_{in}\rangle$, where $R \equiv r_V e^{i\phi_V}|V\rangle\langle V| + r_c(\eta)|H\rangle\langle H|$ and r_V and ϕ_V are respectively the amplitude and phase of the reflection of the reference arm. We choose r_V to match the reflection amplitude of the empty (lossy) cavity, such that, in the absence of an atom, the light emerges in the incident polarization state $|D\rangle \equiv (|V\rangle + \langle H|)/\sqrt{2}$. In the presence of an atom, for $\phi_V = 0$ and $\eta \gg 1$, input light exits the interferometer predominantly with the orthogonal polarization $|A\rangle \equiv (|V\rangle - |H\rangle)/\sqrt{2}$ (Supplementary Information).

Figure 2a demonstrates the optical phase shift arising from an atom coupled to the cavity. A weak, D-polarized probe field is applied at the interferometer input, and the output power in the A and D ports is recorded as a function of the reference phase, ϕ_V. The phase of the reflected light is shifted by $(1.1 \pm 0.1)\pi$ relative to the case with no atom, and the visibilities of the oscillation with ϕ_V are $44\% \pm 2\%$ and $39\% \pm 2\%$ in the A and D ports, respectively. By repeating this measurement for a range of atom–photon detunings, δ, we observe a 2π change in the reflection phase across the atomic resonance (Fig. 2b), in agreement with equation (1). For the data presented, the events where an atom was not present in the cavity (for example because of escape from the trap) were excluded. The remaining contributions to the reduced fringe visibility are imperfect balancing of the interferometer (~5%), atomic saturation effects (~10%), state-changing scattering processes that leave the atom in a different final state and therefore reveal which path
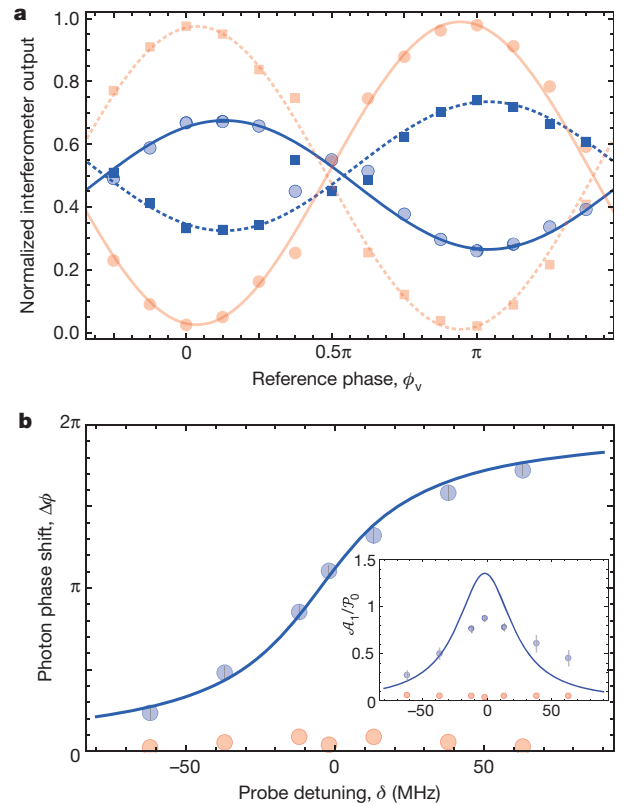

Figure 2 | Photon phase shift produced by a single atom. **a**, Normalized interferometer output versus reference phase, ϕ_V. The blue circles, blue squares, red circles and red squares correspond to $\mathcal{A}_1/\mathcal{P}_1$, $\mathcal{D}_1/\mathcal{P}_1$ (with atom) and $\mathcal{A}_0/\mathcal{P}_0$, $\mathcal{D}_0/\mathcal{P}_0$ (without atom), where $\mathcal{A}$ and $\mathcal{D}$ are the powers in the A and D output ports and $\mathcal{P} \equiv \mathcal{A} + \mathcal{D}$. The measurement is performed near resonance ($\delta = -2$ MHz), and the lines are sinusoidal fits resulting in a phase shift of $(1.1 \pm 0.1)\pi$. The maximum fringe visibilities with and without an atom are $44\% \pm 2\%$ and $97\% \pm 1\%$, respectively. **b**. Measured phase shift versus detuning in the presence (blue) or absence (red) of an atom. The curve includes cavity losses in equation (1) (Supplementary Information) and corresponds to a cooperativity of $\eta = 7.7$ and a small (5-MHz) offset from the free-space resonance. The inset shows $\mathcal{A}_1/\mathcal{P}_0$ at $\phi_V = \pi$. The solid line is the expected value for the same model parameters as in the main figure. The expected increase in reflectivity in the presence of an atom ($\mathcal{P}_1/\mathcal{P}_0 > 1$) arises because the atom reduces the field amplitude in the lossy cavity (Supplementary Information). In our experiment, we observe $\mathcal{P}_1/\mathcal{P}_0 \approx 1.2$. The error bars are the 1σ statistical uncertainty.

the photon has taken in the interferometer (~20%), and thermal motion of the atom (~20%) (Supplementary Information).

The saturation behaviour of the atom–cavity system is examined in Fig. 3a, which shows the fraction of the output power in the A and D ports as a function of the input power. We set the reference phase to $\phi_V \approx 0$, such that the A port is dark in the absence of the atom. The distribution of the output is power independent for low input powers, as expected for a linear system. At higher powers, the atomic response saturates and the output fraction at the A port decreases. The saturation becomes evident when the input photon rate approaches the enhanced excited-state decay rate, Γ, in agreement with theoretical predictions (Supplementary Information). This nonlinearity results in different reflection phases for single photons and photon pairs. In a Hanbury–Brown–Twiss experiment, we measure the photon–photon correlation functions, $g^{(2)}(\tau)$, at low input power (Fig. 3b, c). We observe strong antibunching of $g_A^{(2)}(0) = 0.12(5)$ and bunching of $g_D^{(2)}(0) = 4.1(2)$ at the A and D ports, respectively, indicating that the atom–cavity system acts as an effective photon router by sending single photons to output A and photon pairs to output D (ref. 27).
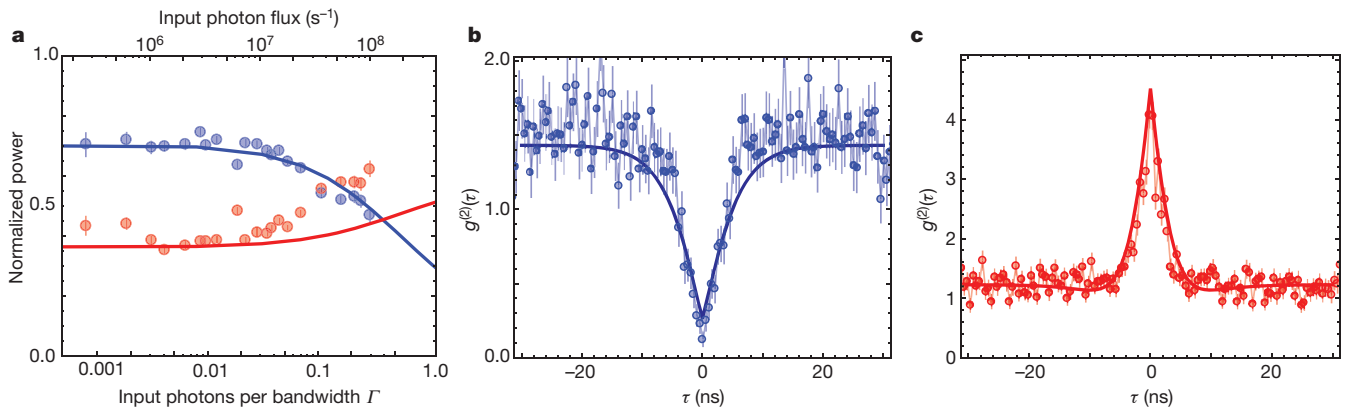
Figure 3 | Quantum nonlinear optics with the atom/photonic-crystal system. a, Interferometer output as a function of the rate at which photons are incident on the interferometer. The outputs $\mathcal{A}_1/\mathcal{P}_0$ (blue) and $\mathcal{D}_1/\mathcal{P}_0$ (red) are normalized to the case without an atom. The incident photon rate is normalized to the enhanced atomic decay rate, $\Gamma = (\eta + 1)\gamma$. The interferometer is tuned such that port A is dark in the absence of the atom and such that the output in port A starts to saturate at a rate less than one photon per bandwidth Γ. Unlike the data in Figs 2 and 4, these measurements were performed in the presence of the dipole trap, which reduces $\mathcal{A}_1/\mathcal{P}_1$ at low driving intensities (Supplementary Information). **b, c**, Photon–photon correlation functions, $g^{(2)}(\tau)$, for the A (**b**) and D (**c**) ports. Port A shows clear antibunching with $g_A^{(2)}(0) = 0.12(5)$, whereas port D exhibits strong bunching with $g_D^{(2)}(0) = 4.1(2)$. The solid lines in **a**–**c** are obtained from a model including inhomogeneous light-shift broadening arising from the dipole trap (Supplementary Information). The error bars are the 1σ statistical uncertainty.

To realize a quantum switch in which the state of a single atom controls the propagation of many probe photons, we use two atomic hyperfine states, $|c\rangle \equiv |F=2, m_F=0\rangle$ and $|u\rangle \equiv |F=1, m_F=0\rangle$ (Fig. 4a), which can be coherently manipulated with microwaves. Although the atom–photon interaction strength is similar for all of the sublevels in a given hyperfine manifold, the $F=1$ levels (including $|u\rangle$) are effectively uncoupled because the probe is far detuned from all optical transitions originating from this level. In Fig. 4a, we show the output signal at the A port for a D-polarized probe field with an atom prepared in the $F=1$ or $F=2$ manifold. The switch is 'on' and the input light goes mostly to the A port when $F=2$, whereas the switch is 'off' and the A port is dark when $F=1$. We estimate that up to $\bar{n}_A \approx 75$ photons could be transmitted to the A port in the 'on' state before the atom is optically pumped out of the $F=2$ manifold. In the experiments shown in Fig. 4, a smaller

number of photons ($\bar{n}_A = 6.2$) was used to increase the rate of data acquisition by allowing a greater number of measurements with the same atom. This photon number allows us to distinguish the switch state with an average fidelity of 95%.

Because the effect of an atom on a photon and that of a photon on an atom are complementary, it follows from equation (1) that a single photon can shift the phase of the coupled state, $|c\rangle$, by π. This phase shift can be converted into a flipping of the atomic switch, $|c\rangle \leftrightarrow |u\rangle$, using an atomic Ramsey interferometer[18]. An atom is first prepared in $|u\rangle$ by means of optical pumping, and then rotated to the superposition $(|u\rangle + |c\rangle)/\sqrt{2}$ by a microwave $\pi/2$-pulse (Supplementary Information). A single H-polarized 'gate' photon flips the atomic superposition to $(|u\rangle - |c\rangle)/\sqrt{2}$. Because reflection of the gate photon does not reveal the atomic state, the atomic superposition is not destroyed.
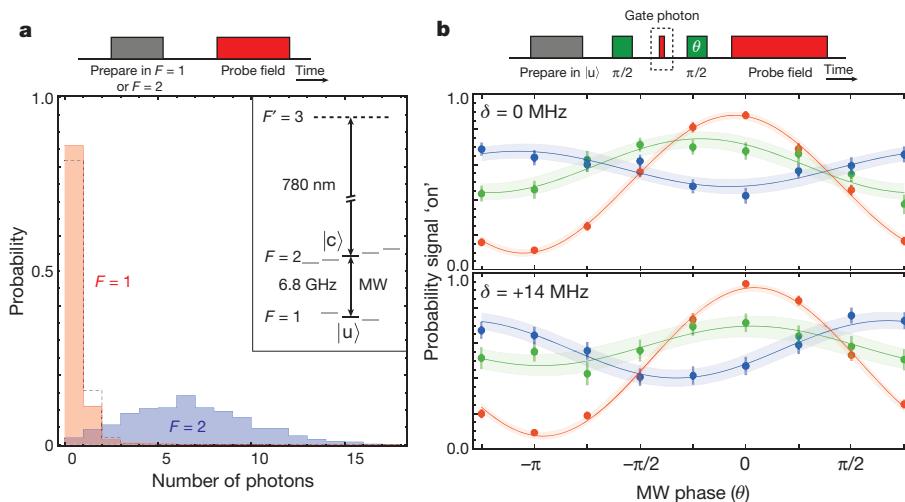

Figure 4 | Realization of the quantum phase switch. a, Number of probe photons detected in port A as a function of the internal atomic state. If the atom is in the $F=2$ manifold, the switch field is 'on', thereby routing $\bar{n}_A = 6.2$ photons to port A (the switch sequence is shown at top). If the atom is absent (dashed line) or in the $F=1$ manifold, then $\bar{n}_A = 0.2$. The input photon number is the same in all cases, with a peak rate much smaller than Γ. The separation between the two distributions allows the switch states to be distinguished with 95% average fidelity. The inset shows the relevant levels for the quantum switch. The laser is tuned to the $F=2$-to-$F'=3$ transition, and

couples only to $|c\rangle$. MW, microwave. **b**, Top: the switch sequence (see text). Bottom: the probability, P_{on}, of finding the switch 'on', as a function of the phase, θ, of the second microwave pulse ($\delta = 0$ (top panel) and $\delta = 2\pi \times 14$ MHz (bottom panel)). P_{on} is shown in several cases: without a gate field (P_{on}^0, red); and with a gate field, both with (P_{on}^1, blue) and without (P_{on}^{uc}, green) conditioning on the detection of a reflected photon. The error bars are the 1σ statistical uncertainty in the data, and the shaded region shows the range of curves with fit parameters within 1σ of the best fit.

Finally, a second microwave $\pi/2$-pulse rotates the atomic state to $|c\rangle$ or $|u\rangle$ depending on the presence or absence of the gate photon, leaving the switch on (atom in $|c\rangle$) or off (atom in $|u\rangle$). A similar technique was recently explored for non-destructive photon detection in a Fabry–Pérot cavity[12].

In our measurement, we mimic the action of a single gate photon by applying a weak coherent field with $\bar{n} \approx 0.6$ incident photons, and measuring the probe transmission conditioned on the detection of a reflected gate photon at either interferometer output. Figure 4b shows the probability, P_{on}, of finding the switch in the 'on' state as a function of the phase of the second microwave pulse. The dependence of P_{on} on the microwave phase when a reflected gate photon is detected shows that the superposition phase is shifted by $(0.98 \pm 0.07)\pi$. The atomic coherence is reduced but not destroyed. The absence of a phase shift in the unconditioned data (Fig. 4b, green curve) confirms that the switch is toggled by a single photon. The phase shift depends on the gate photon detuning: tuning the laser to $\delta = 2\pi \times 14$ MHz results in a phase shift of $(0.63 \pm 0.15)\pi$, in good agreement with the detuning dependence of the photon phase shift (Fig. 2b).

For an optimally chosen phase of the second microwave pulse, we find that the switch is in the 'on' state with probability $P_{on}^1 = 0.64 \pm 0.04$ if a gate photon is detected, $P_{on}^0 = 0.11 \pm 0.01$ if no gate field is applied and $P_{on}^{uc} = 0.46 \pm 0.06$ if we do not condition on single-photon detection. The finite $P_{on}^0 > 0$ without a gate field arises from imperfect atomic-state preparation and readout fidelity (Supplementary Information). Also, P_{on}^1 is affected by the finite probability of the gate field containing two photons, of which only one is detected. This results in a decrease in P_{on}^1 and an increase in P_{on}^{uc} by about 20% in a way that is consistent with our measurements (Methods Summary and Supplementary Information). We attribute the 8% positive offset in P_{on}^1 and P_{on}^{uc} to spontaneous scattering events of the gate photon, which cause atomic transitions to a final state other than $|c\rangle$ within the $F = 2$ manifold. Lastly, we estimate that fluctuations in η arising from thermal motion do not change P_{on}^1 by more than 10%, because the atom–photon interaction scheme used here[8] is inherently robust to variations in η for $\eta \gg 1$. The imperfect fringe visibility in Figs 2 and 4, which is due to the technical imperfections discussed above, can be improved by better atomic-state preparation, alignment of the cavity polarization with the magnetic field defining the quantization axis, and improved atom localization. The fringe visibility does not directly depend on the cooperativity and, absent technical imperfections, perfect fringe visibility should be achievable; however, the probability of gate photon loss is reduced as the cooperativity increases (Supplementary Information).

Our experiments enable a number of intriguing applications. For instance, efficient atom–photon entanglement for quantum networks can be generated by reflecting a single photon from an atom prepared in a superposition state. The quantum phase switch also allows for quantum non-demolition measurements of optical photons[12,28]. With an improved collection efficiency of light from the photonic crystal cavity and reduced cavity losses, it should be possible to make high-fidelity non-demolition measurements of optical photon number parity to create non-classical Schrödinger cat states[29], with possible applications to state purification and error correction. Most notably, the scalable nature of both nanofabrication and atomic trapping allow for extensions of this work to complex integrated networks with multiple atoms and photons.

METHODS SUMMARY

We begin our experiments by loading a single ^{87}Rb atom from a magneto-optical trap into a tightly focused optical dipole trap. After a period of Raman sideband cooling[30] to localize the atom in the trapping potential, we translate the optical dipole trap to the photonic crystal cavity, where the interference of the dipole trap light with its reflection from the photonic crystal forms an intensity maximum that confines the atom at a distance of about 200 nm from the surface of the photonic crystal[25] (Fig. 1a, b). The success probability of loading an atom near the photonic crystal is >90%. We modulate the dipole trap with full contrast at 5 MHz to interrogate the trapped atom at instances over which the light shift is negligible.

1. Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* **78,** 3221–3224 (1997).
2. Kimble, H. J. The quantum internet. *Nature* **453,** 1023–1030 (2008).
3. Duan, L.-M. & Monroe, C. Quantum networks with trapped ions. *Rev. Mod. Phys.* **82,** 1209–1224 (2010).
4. Haroche, S. & Raimond, J.-M. *Exploring the Quantum: Atoms, Cavities, and Photons* (Oxford Univ. Press, 2006).
5. Briegel, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81,** 5932–5935 (1998).
6. Kómár, P. *et al.* A quantum network of clocks. Preprint at http://arxiv.org/abs/1310.6045 (2013).
7. Carusotto, I. & Ciuti, C. Quantum fluids of light. *Rev. Mod. Phys.* **85,** 299–366 (2013).
8. Duan, L.-M. & Kimble, H. J. Scalable photonic quantum computation through cavity-assisted interactions. *Phys. Rev. Lett.* **92,** 127902 (2004).
9. Schuster, I. *et al.* Nonlinear spectroscopy of photons bound to one atom. *Nature Phys.* **4,** 382–385 (2008).
10. Aoki, T. *et al.* Efficient routing of single photons by one atom and a microtoroidal cavity. *Phys. Rev. Lett.* **102,** 083601 (2009).
11. Chen, W. *et al.* All-optical switch and transistor gated by one stored photon. *Science* **341,** 768–770 (2013).
12. Reiserer, A., Ritter, S. & Rempe, G. Nondestructive detection of an optical photon. *Science* **342,** 1349–1351 (2013).
13. O'Shea, D., Junge, C., Volz, J. & Rauschenbeutel, A. Fiber-optical switch controlled by a single atom. *Phys. Rev. Lett.* **111,** 193601 (2013).
14. Volz, T. *et al.* Ultrafast all-optical switching by single photons. *Nature Photon.* **6,** 605–609 (2012).
15. Kim, H., Bose, R., Shen, T. C., Solomon, G. S. & Waks, E. A quantum logic gate between a solid-state quantum bit and a photon. *Nature Photon.* **7,** 373–377 (2013).
16. Chang, D. E., Sorensen, A. S., Demler, E. A. & Lukin, M. D. A single-photon transistor using nanoscale surface plasmons. *Nature Phys.* **3,** 807–812 (2007).
17. Schuster, D. I. *et al.* Resolving photon number states in a superconducting circuit. *Nature* **445,** 515–518 (2007).
18. Gleyzes, S. *et al.* Quantum jumps of light recording the birth and death of a photon in a cavity. *Nature* **446,** 297–300 (2007).
19. Deléglise, S. *et al.* Reconstruction of non-classical cavity field states with snapshots of their decoherence. *Nature* **455,** 510–514 (2008).
20. Turchette, Q. A., Hood, C. J., Lange, W., Mabuchi, H. & Kimble, H. J. Measurement of conditional phase shifts for quantum logic. *Phys. Rev. Lett.* **75,** 4710–4713 (1995).
21. Fushman, I. *et al.* Controlled phase shifts with a single quantum dot. *Science* **320,** 769–772 (2008).
22. Aoki, T. *et al.* Observation of strong coupling between one atom and a monolithic microresonator. *Nature* **443,** 671–674 (2006).
23. Ritter, S. *et al.* An elementary quantum network of single atoms in optical cavities. *Nature* **484,** 195–200 (2012).
24. Devoret, M. H. & Schoelkopf, R. J. Superconducting circuits for quantum information: an outlook. *Science* **339,** 1169–1174 (2013).
25. Thompson, J. D. *et al.* Coupling a single trapped atom to a nanoscale optical cavity. *Science* **340,** 1202–1205 (2013).
26. Waks, E. & Vuckovic, J. Dispersive properties and large Kerr nonlinearities using dipole-induced transparency in a single-sided cavity. *Phys. Rev. A* **73,** 041803 (2006).
27. Witthaut, D., Lukin, M. D. & Sörensen, A. S. Photon sorters and QND detectors using single photon emitters. *Europhys. Lett.* **97,** 50007 (2012).
28. Volz, J., Gehr, R., Dubois, G., Esteve, J. & Reichel, J. Measurement of the internal state of a single atom without energy exchange. *Nature* **475,** 210–213 (2011).
29. Wang, B. & Duan, L.-M. Engineering superpositions of coherent states in coherent optical pulses through cavity-assisted interaction. *Phys. Rev. A* **72,** 022320 (2005).
30. Thompson, J. D., Tiecke, T. G., Zibrov, A. S., Vuletić, V. & Lukin, M. D. Coherence and Raman sideband cooling of a single atom in an optical tweezer. *Phys. Rev. Lett.* **110,** 133001 (2013).

Author Contributions The experiments and analysis were carried out by T.G.T., J.D.T., N.P.d.L. and L.R.L. All work was supervised by V.V. and M.D.L. All authors discussed the results and contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.D.L. (lukin@fas.harvard.edu) or V.V. (vuletic@mit.edu).

LETTER

Dynamics of continental accretion

L. Moresi[1,2,3], P. G. Betts[1], M. S. Miller[4] & R. A. Cayley[5]

Subduction zones become congested when they try to consume buoyant, exotic crust. The accretionary mountain belts (orogens) that form at these convergent plate margins have been the principal sites of lateral continental growth through Earth's history. Modern examples of accretionary margins are the North American Cordilleras and southwest Pacific subduction zones. The geologic record contains abundant accretionary orogens, such as the Tasmanides[1], along the eastern margin of the supercontinent Gondwana, and the Altaïdes, which formed on the southern margin of Laurasia[2]. In modern and ancient examples of long-lived accretionary orogens, the overriding plate is subjected to episodes of crustal extension and back-arc basin development, often related to subduction rollback[3] and transient episodes of orogenesis and crustal shortening[4–7], coincident with accretion of exotic crust. Here we present three-dimensional dynamic models that show how accretionary margins evolve from the initial collision, through a period of plate margin instability, to re-establishment of a stable convergent margin. The models illustrate how significant curvature of the orogenic system develops, as well as the mechanism for tectonic escape of the back-arc region. The complexity of the morphology and the evolution of the system are caused by lateral rollback of a tightly arcuate trench migrating parallel to the plate boundary and orthogonally to the convergence direction. We find geological and geophysical evidence for this process in the Tasmanides of eastern Australia, and infer that this is a recurrent and global phenomenon.

Since the origination of the concept of suspect terranes[8], accretionary orogenic belts have been identified in the continent record at every stage in Earth's history[9]. Accretionary orogenic belts form at intra-oceanic and continental convergent plate boundaries and comprise supra-subduction fore-arc, arc and back-arc regions, as well as microcontinents that are caught within the belts. They remain active over long geologic timescales (tens to hundreds of millions of years), and affect a broad range of scientific disciplines because they are the sites of significant variation in continental and sea-floor topography, which influences global climate patterns, ocean circulation and biota. From a geologic perspective, accretionary orogenic systems are sites of significant lateral continental growth via the accretion of juvenile crust and allochthonous microcontinents onto plate margins, and they drive plate reorganization and cratonization[7,9]. Implicit in the evolution of accretionary orogens is the transfer of buoyant crustal material from the downgoing plate to the overriding plate and the re-establishment of subduction to accommodate continued plate convergence[10]. How and where subduction re-establishes once a terrane has been accreted remains a fundamental and significant unresolved question related to accretionary orogenic belt evolution and plate tectonic theory.

We address the evolution of accretionary orogens in general by re-assessing the evolution of the southern Tasmanides[11] with numerical models. The Tasmanides, which represent the Australian segment of the larger Terra Australis orogen[12], were formed by a long-lived accretionary phase of the convergent margin between the Palaeo-Pacific plate and East Gondwana between the late Neoproterozoic era and the Jurassic period, and is arguably still evolving today[13]. This accretionary belt is responsible for the growth of approximately one-third of

the Australian continent[11] with the present-day subduction zones of the southwest Pacific representing the continued evolution of this accretionary belt. The Tasmanides comprise multiple accreted microcontinents, arcs and back-arc terranes, which now form a collage of contorted geologic belts[13] that extend the entire length of the Australian continent (Fig. 1a). It represents the 'type' setting for a retreating accretionary orogen punctuated by transient episodes of crustal shortening and compressional orogenesis[7]. Major variation in the geometry and evolution of the accretion are documented along the strike-length of the belt and consequently there is little consensus concerning the tectonic setting and orogen drivers, polarity of subduction zones, and the role of continental ribbons and subduction rollback in the evolution (see below).

In order to better understand the behaviour of this ancient plate margin and the growth of the Australian continent, we use three-dimensional (3D) dynamic models of a subducting slab, overriding plate and mantle, building on previous work[14–16]. The models have a four-layer subducting plate with buoyancy and rheology of each layer pre-calculated from a half-space cooling model of 80 or 120 Myr age, and they include either a weak or a strong viscoplastic overriding plate (Table 1 and Extended Data Figs 1 and 2). The simulations are best understood by viewing movies of the time evolution (Table 1).

All four models show the same three stages of evolution: (1) a collisional stage when the microcontinental ribbon initially accretes to the overriding plate; (2) a transitional stage where the convergent subducting plate and trench reorganize through coeval trench advance and retreat in different parts of the boundary; and (3) the re-initiation of a stable subduction system behind the accreted microcontinent. We describe this in detail using model 80s which has an 80-Myr-old lithosphere and a strong overriding plate. The comparative evolution of models 80w and 120s are shown in Extended Data Figs 3 and 4.

At the onset of the collision stage (Figs 2a and 3), the plate boundary develops two distinct domains: a region of distributed shortening along the length of the trench that is congested by the microcontinent, and a region of extension associated with the continued retreat of the active subduction zone. The tip of the microcontinent is at the transition between these two domains and this is where strike-slip deformation perpendicular to the margin is concentrated. Subduction in the congested region of the trench rapidly stalls and the slab begins to stretch, creating a window or tear in the slab. Distal to the colliding ribbon, slab rollback and trench retreat continue, resulting in trench-perpendicular extension and rifting of the overriding plate. In the overriding plate ahead of the colliding ribbon, material begins to extrude laterally into the region of rollback-driven crustal extension.

During the second stage of the model (Fig. 2b), the retreating trench reaches the old oceanic lithosphere at the rear of the continental ribbon. The trench is pinned at the rearward tip of the ribbon and begins a lateral retreat behind the accreted terrane. This results in a tightly curved arc, a synclinal slab geometry, and material from the overriding plate being dragged behind the ribbon (Fig. 2c and Extended Data Fig. 5). The slab attached to the margin of the continental ribbon is emplaced by a foundering of the oceanic lithosphere beneath the rifted and extended overriding plate followed by a progressive steepening (Extended Data

[1]School of Geosciences, Monash University, Clayton, Victoria 3800, Australia. [2]School of Mathematical Sciences, Monash University, Clayton, Victoria 3800, Australia. [3]School of Earth Sciences, University of Melbourne, Parkville, Victoria 3010, Australia. [4]Department of Earth Sciences, University of Southern California, Los Angeles, California 90089, USA. [5]Geological Survey of Victoria, Melbourne, Victoria 3001, Australia.
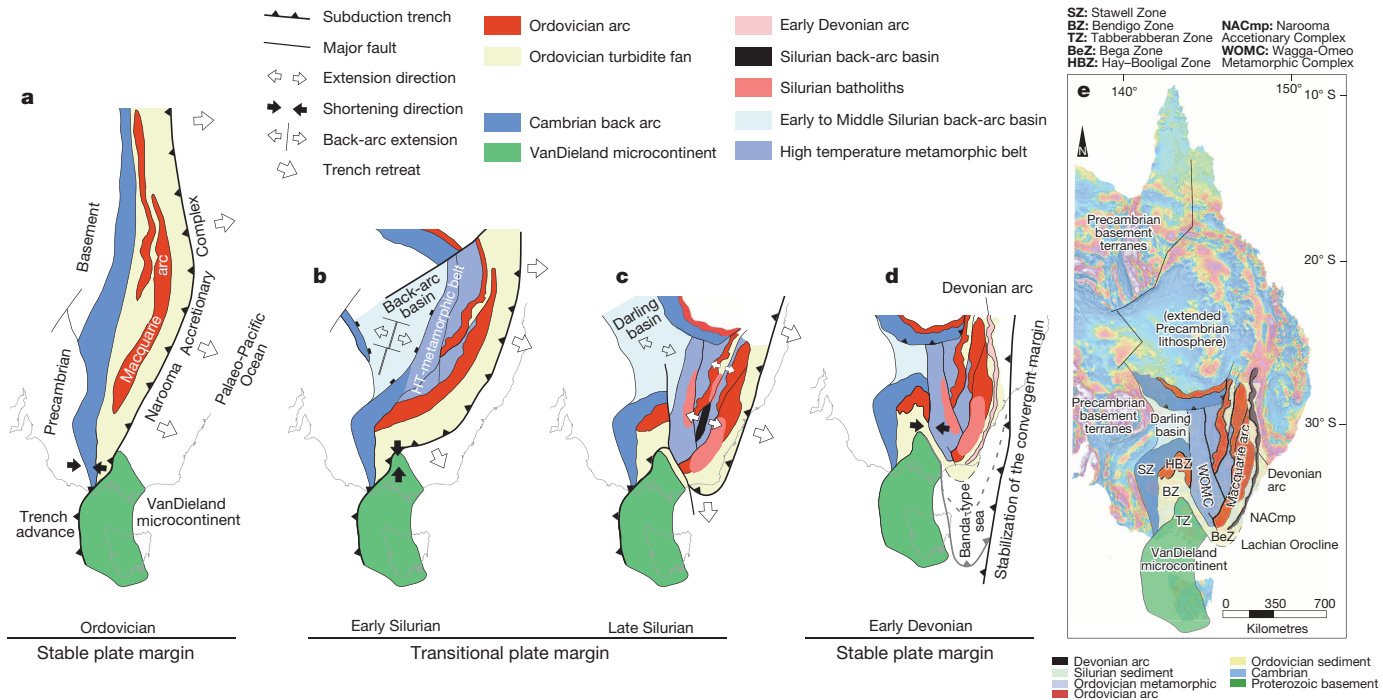
Figure 1 | Reconstruction of the Tasmanide accretionary event, showing indentation of continental ribbon material and lateral subduction. a, Stable Ordovician convergent margin immediately after accretion of VanDieland microcontinent. **b,** Trench advance in the southern Tasmanides and subduction rollback in the central Tasmanides during the Early Silurian resulted in a transitional configuration of back-arc extension, and high temperature metamorphism in the back arc. **c,** Margin parallel trench retreat behind VanDieland during the Late Silurian. Extensive arc and back-arc magmatism, and basins systems above the retreating plate leading to a Banda-type sea outboard of VanDieland. **d,** Re-establishment of a stable, linear convergent margin during the Early Devonian. **e,** Present distribution of major tectonic elements of the southern and central Tasmanides superimposed on a magnetic map of eastern Australia (modified after ref. 22 using data available from Geoscience Australia at the Geophysical Archive Data Delivery System, http://www.geoscience.gov.au/gadds).

Fig. 6). The slab outboard of the margin accommodates the convergence by continued rollback subduction. The laterally retreating part of the slab undergoes pure trench rollback subduction at a rate comparable to the convergence velocity of the two plates (Fig. 3). The boundary between shortening and extension in the overriding plate is where deformation is concentrated and is associated with the contact point between the retreating trench and the rear margin of the ribbon. The progression of this point from the tip of the ribbon to its interior creates the orocline. In models with a strong overriding plate, a switch to extension in the whole of the region ahead of the ribbon allows rapid development of localized shear zones. At this point, rates of tectonic escape from behind the indenting ribbon are greatly enhanced through the formation of a major, sinistral, strike-slip structure (Fig. 2c).

At the end of this transitional phase, indentation of the ribbon is complete, subduction is re-established outboard of the orocline (Fig. 2d), and the overriding plate behind the ribbon returns to a mildly extensional setting (Fig. 3). The accreted terrane is deeply embedded in the overriding plate and separated from the active margin by a zone of attenuated crust derived from the overriding plate and back-arc spreading associated with lateral rollback. The geometry of the convergent margin is relatively linear and is again oriented perpendicular to the convergence direction. Subduction has resumed significantly outboard of the accreted terrane and orocline. In these models, the overall convergence direction remains uniform due to the domain boundary. There is no subduction polarity reversal, but lateral rollback means that, in many cross-sections, slabs appear to be dipping in opposite directions, as is often interpreted in tectonic reconstructions of the Tasmanides[17].

For model 80s, shown in Fig. 3, the initial collision is accompanied by a slowing of the oceanic plate velocity, a switch of the congested section of the plate boundary to advance, and subduction of the leading edge of the ribbon (7 Myr to 14 Myr from the start of the model run). After the suture forms, the retreat rate of the uncongested trench reaches a peak, before reducing as the lateral rollback of the slab begins (24 Myr) and decreasing in parallel with the increasing rate of lateral rollback. At

Table 1 | Parameters and Supplementary Videos for models 80w, 80s, 120w and 120s

Model name	Movie	View*	Age of oceanic lithosphere (Myr)	Cohesion of back-arc region crust (MPa)	Post-softening cohesion of back-arc region crust (MPa)	Cohesion of back-arc region lithosphere (MPa)	Post-softening cohesion of back-arc region lithosphere (MPa)
80w	Supplementary Video 3	1	80	85	42.5	170	85
80s	Supplementary Video 4	1	80	170	85	340	170
80s	Supplementary Video 1	2	80	170	85	340	170
120w	Supplementary Video 5	1	120	85	42.5	170	85
120s	Supplementary Video 6	1	120	170	85	340	170
120s	Supplementary Video 2	2	120	170	85	340	170

* View 1, looking along the line of the original trench location from above; view 2, looking at the underside of the collision from below the overriding plate. The white/red balls in the second view are finite strain markers located under the oceanic plate at the start of the simulation. The strain rate field close to the surface is superimposed on the colouring of the material domains to show high strain rates as red tones and intermediate strain rates in yellow/green.
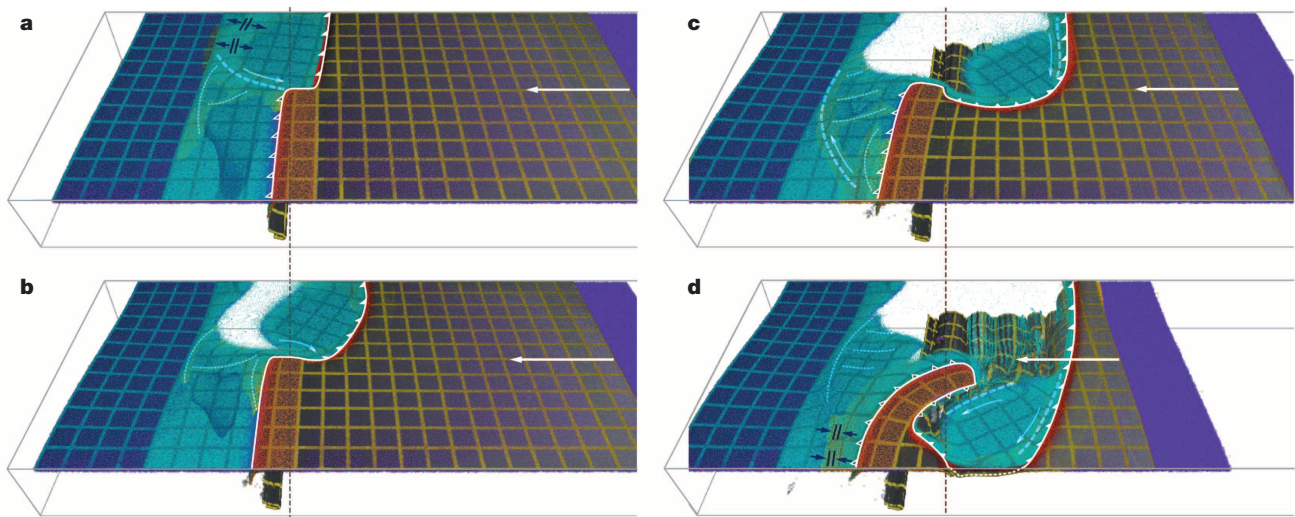
Figure 2 | Four frames in the evolution of model 80s taken at 18, 27, 39 and 61 Myr from the start of the simulation. Time in Myr from start of simulation: **a**, 18; **b**, 27; **c**, 39; and **d**, 61. The oceanic plate is grey with a yellow strain-marker grid. The continental crust is dark blue with a light grid. Transitional crust is in light blue. The indenting ribbon is red/orange. The plate boundary is a white line: solid teeth show active subduction, open teeth indicate underthrusting and suturing. The dotted region indicates the foundering of the plate. Shear-strain localization is marked by dashed lines, those with largest deformation are heavy with the sense of motion indicated. Regions of tectonic compression and extension are indicated by dark arrows. The dark, dashed line indicates the position of the trench at the onset of collision (7 Myr).

40 Myr, a diffuse plate boundary develops in the oceanic plate and the indentation of the ribbon slows rapidly while the oceanic plate sub-duction rate increases. For model 120s (Extended Data Fig. 4), this pattern is very similar but the lateral migration of the subduction zone is slower and the diffuse plate boundary takes longer to develop. Model 80w, which has a weaker overriding plate, evolves more rapidly fol-lowing a very similar pattern (Extended Data Fig. 4).

In models with an 80-Myr-old slab, the formation of a slab window occurs when the suture forms and at the same time as the onset of rapid rollback associated with extension in the overriding plate. Mantle flow through the slab window (Supplementary Video 1) has a strong toroidal component familiar from other studies[14] which sweeps material from behind the original trench location into the back-arc region of lateral rollback. By contrast, the older slab in model 120s stretches and forms a window only after significant slab advance (Supplementary Video 2). This coincides with a jump in the rate of shortening of the overriding

plate at 25 Myr from the start of the model (Extended Data Fig. 4). In this model, none of the marked material from the oceanic mantle domain reaches the laterally retreating back-arc region.

Our computational models compare very well with recent kinematic models for the pre-Carboniferous evolution of the southern Tasmanides of eastern Australia[18]. In the Ordovician to Early Silurian, the VanDieland microcontinent[19], an amalgam of Mesoproterozoic to Cambrian con-tinental, oceanic and arc terranes[18], accreted onto the edge of the east Gondwana margin. This triggered an episode of crustal shortening in the southern Tasmanides (Fig. 1a).

Rollback of the slab to the north of the microcontinent termination resulted in development of the Darling and Newell basins in the back-arc region[20] (Fig. 1b, c). In particular, the Darling basin formed in a position comparable to that predicted by our models and coincides with thinned continental crust[21]. Margin parallel lateral propagation of the Silurian subduction trench resulted in transfer of Ordovician
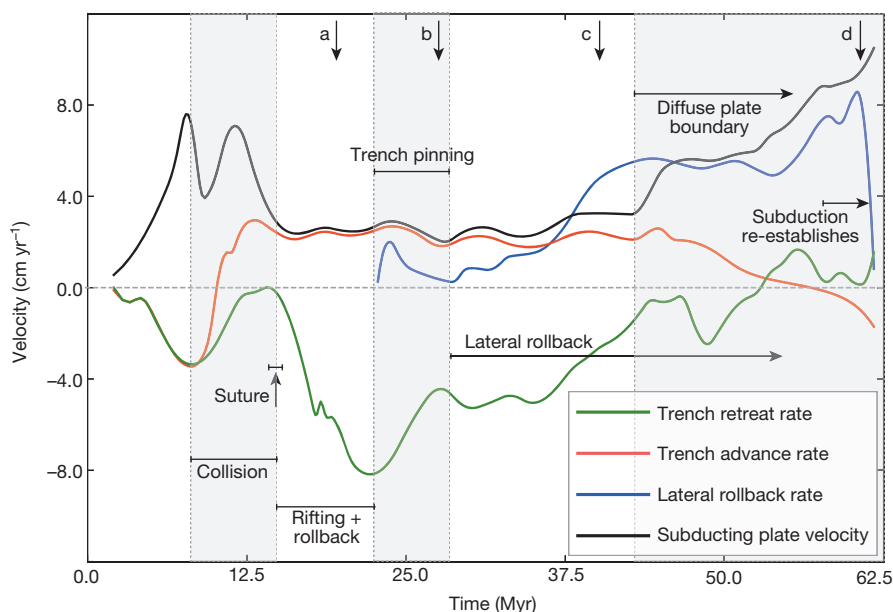

Figure 3 | A comparison of velocities during collision, accretion and recovery. Shown are the velocities of the subducting plate (black lines), the retreating end of the plate boundary ('Trench retreat rate', green), the colliding end of the plate boundary ('Trench advance rate', red), and the cusp of the laterally retreating trench ('Lateral rollback rate', blue). Negative velocities indicate retreat, positive velocities indicate advance. The markers (a–d) indicate the times of the frames in Fig. 2.

arc terranes and back-arc turbidites from the overriding plate to the overriding plate behind the accreted microcontinent, as seen in our geodynamic models. The Lachlan orocline (Fig. 1e) formed behind VanDieland and is defined by arcuate belts of Ordovician turbidites and Silurian arc and back-arc magmas[22] (Fig. 1b and c). Re-establishment of a stable linear subduction behind the accreted VanDieland is defined by the emplacement of latest Silurian to Early Devonian arc magmas, which define a linear north-trending belt (Fig. 1d)

This dynamical model of the pre-Carboniferous evolution of the southern Tasmanides highlights the importance of the 3D subduction dynamics at the tip of a ribbon microcontinent. Such locations are the sites of transition from collision-related compression to collision-driven rollback and extension in the overriding plate. This transition creates the shear that drives the formation of an orocline. Slabs re-organize and migrate behind accreted terranes in a characteristic fashion that draws material from the overriding plate into the collision zone. This implies that all such collisions need to be reinterpreted.

A feature of all of our models is the emergence of small-scale, fast moving, laterally migrating subduction zones typical of late-stage ocean closure such as the Calabrian, Gibraltar and Hellenic arcs of the Mediterranean[23,24]. Our models also show the partitioning of compression, extension and block rotation common in collisional margins[25], and demonstrate the characteristics of the long-lived cycle of subduction with repeated collisions observed in the northern Cordillera of Alaska and western Canada[26]. The same mechanism is also seen at a much larger scale in the convergence of India with Eurasia, namely tectonic escape coupled with regional rollback that is seen in the combined Himalayan indentation[27,28], and the retreat with further extension seen in the South East Asia arcs of Sumatra, Java and Banda[29]. Our models can explain both geologically recent processes[23–29] and some ancient orogenesis, such as that which occurred during the Phanerozoic in Australia and Central Asia, and during the Proterozoic in the Americas and Europe[2,9].

METHODS SUMMARY

We use 3D numerical dynamic models of a 7,000 km freely subducting slab, overriding plate and mantle building on previous work[14–16] using the same numerical implementation of the particle-in-cell finite element method in the code UNDERWORLD[30]. Models consist of a 3D Cartesian domain (Extended Data Fig. 1) 800 km deep and extending 8,000 × 3,000 km horizontally in which we solve the Stokes equation that balances buoyancy forces with viscous stresses and pressure gradients. We assume incompressibility and neglect thermal diffusion. Material properties are carried by Lagrangian integration points and mapped to element properties. This numerical approach is designed to accurately track distinct material domains in the lithosphere as they undergo large deformation due to creeping flow.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Glen, R. A. The Tasmanides of eastern Australia. *Geol. Soc. Lond. Spec. Publ.* **246**, 23–96 (2005).
2. Şengör, A. M. C., Natal'in, B. A. & Burtman, V. S. Evolution of the Altaid tectonic collage and Paleozoic crustal growth in Asia. *Nature* **364**, 299–307 (1993).
3. Wilhem, C., Windley, B. F. & Stampfli, G. M. The Altaids of Central Asia: a tectonic and evolutionary innovative review. *Earth Sci. Rev.* **113**, 303–341 (2012).
4. Dewey, J. F., Hempton, M. R., Kidd, W. S. F., Saroglu, F. & Şengör, A. M. C. Shortening of continental lithosphere: the neotectonics of Eastern Anatolia a young collision zone. *Geol. Soc. Lond. Spec. Publ.* **19**, 1–36 (1986).
5. Lister, G. S., Forster, M. A. & Rawling, T. J. Episodicity during orogenesis. *Geol. Soc. Lond. Spec. Publ.* **184**, 89–113 (2001).
6. Giles, D., Betts, P. & Lister, G. Far-field continental backarc setting for the 1.80–1.67 Ga basins of northeastern Australia. *Geology* **30**, 823–826 (2002).
7. Collins, W. J. Hot orogens, tectonic switching, and creation of continental crust. *Geology* **30**, 535–538 (2002).
8. Coney, P. J., Jones, D. L. & Monger, J. W. H. Cordilleran suspect terranes. *Nature* **288**, 329–333 (1980).
9. Cawood, P. A. *et al.* Accretionary orogens through Earth history. *Geol. Soc. Lond. Spec. Publ.* **318**, 1–36 (2009).
10. Baes, M., Govers, R. & Wortel, M. J. R. Subduction initiation along the inherited weakness zone at the edge of a slab: insights from numerical models. *Geophys. J. Int.* **184**, 991–1008 (2011).
11. Glen, R. A. Refining accretionary orogen models for the Tasmanides of eastern Australia. *Aust. J. Earth Sci.* **60**, 315–370 (2013).
12. Cawood, P. A. Terra Australis Orogen: Rodinia breakup and development of the Pacific and Iapetus margins of Gondwana during the Neoproterozoic and Paleozoic. *Earth Sci. Rev.* **69**, 249–279 (2005).
13. Glen, R. A. & Meffre, S. Styles of Cenozoic collisions in the western and southwestern Pacific and their applications to Palaeozoic collisions in the Tasmanides of eastern Australia. *Tectonophysics* **479**, 130–149 (2009).
14. Schellart, W., Stegman, D., Farrington, R., Freeman, J. & Moresi, L. Cenozoic tectonics of western North America controlled by evolving width of Farallon slab. *Science* **329**, 316–319 (2010).
15. Mason, W. G., Moresi, L., Betts, P. G. & Miller, M. S. Three-dimensional numerical models of the influence of a buoyant oceanic plateau on subduction zones. *Tectonophysics* **483**, 71–79 (2010).
16. Betts, P. G., Mason, W. G. & Moresi, L. The influence of a mantle plume head on the dynamics of a retreating subduction zone. *Geology* **40**, 739–742 (2012).
17. Gray, D. R. & Foster, D. A. Orogenic concepts — application and definition: Lachlan fold belt, eastern Australia. *Am. J. Sci.* **297**, 859–891 (1997).
18. Moore, D. H., Betts, P. G. & Hall, M. Towards understanding the early Gondwanan margin in southeastern Australia. *Gondwana Res.* **23**, 1581–1598 (2013).
19. Cayley, R. A. Exotic crustal block accretion to the eastern Gondwanaland margin in the Late Cambrian-Tasmania, the Selwyn Block, and implications for the Cambrian-Silurian evolution of the Ross, Delamerian, and Lachlan orogens. *Gondwana Res.* **19**, 628–649 (2011).
20. Fergusson, C. L. Plate-driven extension and convergence along the East Gondwana active margin: Late Silurian-Middle Devonian tectonics of the Lachlan Fold Belt, southeastern Australia. *Aust. J. Earth Sci.* **57**, 627–649 (2010).
21. Kennett, B. L. N., Salmon, M., Saygin, E. & AusMoho Working Group. AusMoho: the variation of Moho depth in Australia. *Geophys. J. Int.* **187**, 946–958 (2011).
22. Cayley, R. A. in *The Great East-Australian Arm-Wave of 2012: Tectonics, Modern Analogues/Events, Mineralisation (GAGA 2012)* (ed. White, S.) 34–43 (Selwyn Symposium of the GSA Victoria Division, Geological Society of Australia Extended Abstract 103, 2012).
23. Royden, L. H. Evolution of retreating subduction boundaries formed during continental collision. *Tectonics* **12**, 629–638 (1993).
24. Govers, R. & Wortel, M. J. R. Lithosphere tearing at STEP faults: response to edges of subduction zones. *Earth Planet. Sci. Lett.* **236**, 505–523 (2005).
25. Wallace, L. M., Ellis, S. & Mann, P. Collisional model for rapid fore-arc block rotations, arc curvature, and episodic back-arc rifting in subduction settings. *Geochem. Geophys. Geosyst.* **10**, Q05001 http://dx.doi.org/10.1029/2008GC002220 (2009).
26. Johnston, S. T. The Great Alaskan Terrane Wreck: reconciliation of paleomagnetic and geological data in the northern Cordillera. *Earth Planet. Sci. Lett.* **193**, 259–272 (2001).
27. Molnar, P. & Tapponnier, P. Cenozoic tectonics of Asia: effects of a continental collision. *Science* **189**, 419–426 (1975).
28. England, P. & Houseman, G. Extension during continental convergence, with application to the Tibetan Plateau. *J. Geophys. Res.* **94**, 17561–17579 (1989).
29. Hall, R. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *J. Asian Earth Sci.* **20**, 353–431 (2002).
30. Moresi, L. *et al.* Computational approaches to studying non-linear dynamics of the crust and mantle. *Phys. Earth Planet. Inter.* **163**, 69–82 (2007).

Author Contributions All authors contributed to the conception and writing of the paper and to the geodynamic analysis. L.M. and M.S.M. contributed the geophysical synthesis for the model set-up. P.G.B. and R.A.C. contributed the geological synthesis of the Tasmanides. L.M. ran the numerical models and made the animations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.M. (louis.moresi@unimelb.edu.au).

Detection and replication of epistasis influencing transcription in humans

Gibran Hemani[1,2], Konstantin Shakhbazov[1,2], Harm-Jan Westra[3], Tonu Esko[4,5,6], Anjali K. Henders[7], Allan F. McRae[1,2], Jian Yang[1], Greg Gibson[8], Nicholas G. Martin[7], Andres Metspalu[4], Lude Franke[3], Grant W. Montgomery[7]*, Peter M. Visscher[1,2]* & Joseph E. Powell[1,2]*

Epistasis is the phenomenon whereby one polymorphism's effect on a trait depends on other polymorphisms present in the genome. The extent to which epistasis influences complex traits[1] and contributes to their variation[2,3] is a fundamental question in evolution and human genetics. Although often demonstrated in artificial gene manipulation studies in model organisms[4,5], and some examples have been reported in other species[6], few examples exist for epistasis among natural polymorphisms in human traits[7,8]. Its absence from empirical findings may simply be due to low incidence in the genetic control of complex traits[2,3], but an alternative view is that it has previously been too technically challenging to detect owing to statistical and computational issues[9]. Here we show, using advanced computation[10] and a gene expression study design, that many instances of epistasis are found between common single nucleotide polymorphisms (SNPs). In a cohort of 846 individuals with 7,339 gene expression levels measured in peripheral blood, we found 501 significant pairwise interactions between common SNPs influencing the expression of 238 genes ($P < 2.91 \times 10^{-16}$). Replication of these interactions in two independent data sets[11,12] showed both concordance of direction of epistatic effects ($P = 5.56 \times 10^{-31}$) and enrichment of interaction P values, with 30 being significant at a conservative threshold of $P < 9.98 \times 10^{-5}$. Forty-four of the genetic interactions are located within 5 megabases of regions of known physical chromosome interactions[13] ($P = 1.8 \times 10^{-10}$). Epistatic networks of three SNPs or more influence the expression levels of 129 genes, whereby one *cis*-acting SNP is modulated by several *trans*-acting SNPs. For example, *MBNL1* is influenced by an additive effect at rs13069559, which itself is masked by *trans*-SNPs on 14 different chromosomes, with nearly identical genotype–phenotype maps for each *cis*–*trans* interaction. This study presents the first evidence, to our knowledge, for many instances of segregating common polymorphisms interacting to influence human traits.

In the genetic analysis of complex traits it is usual for SNP effects to be estimated using an additive model in which they are assumed to contribute independently and cumulatively to the mean of a trait. This framework has been successful in identifying thousands of associations[14]. But to date, although its contribution to phenotypic variance is frequently the subject of debate[1–3], there is little empirical exploration of the role that epistasis has in the architecture of complex traits in humans[7,8]. Beyond the prism of human association studies there is evidence for epistasis, not only at the molecular scale from artificially induced mutations[4] but also at the evolutionary scale in fitness adaptation[15] and speciation[16].

Methods are now available to overcome the computational problems involved in searching for epistasis, but its detection still remains problematic owing to reduced statistical power. For example, increased dependence on linkage disequilibrium between causal SNPs and observed SNPs[17,18],

increased model complexity in fitting interaction terms[19], and more extreme significance thresholds to account for increased multiple testing[9] all make it more difficult to detect epistasis in comparison to additive effects. Thus, with small genetic effect sizes, as is expected in most complex traits of interest[14], the power to detect epistasis diminishes rapidly. There are two simple ways to overcome this problem. One is by using extremely large sample sizes[20]; another is by analysing traits that are likely to have large effect sizes among common variants. Because our focus was to ascertain the extent to which instances of epistasis arises from natural genetic variation, we designed a study around the latter approach and searched for epistatic genetic effects that influence gene expression levels. Transcription levels can be measured for thousands of genes and like most complex diseases, these expression traits are typically heritable[21]. But unlike complex diseases, genetic associations with gene expression commonly have very large effect sizes that explain large proportions of the genetic variance[22], making them good candidates to search for epistasis, should it exist.

In our discovery data set (Brisbane Systems Genetics Study, BSGS[23]) of 846 individuals genotyped at 528,509 SNPs, we used a two-stage approach to identify genetic interactions. First, we exhaustively tested every pair of SNPs for pairwise effects against each of 7,339 expression traits in peripheral blood (1.03×10^{15} statistical tests, family-wise error rate of 5% corresponding to a significance threshold of $P < 2.91 \times 10^{-16}$, Methods). Second, we filtered the SNP pairs from stage 1 on linkage disequilibrium and genotype class counts, and tested the remaining pairwise effects for significant interaction terms and used a Bonferroni correction for multiple testing (estimated type 1 error rate $0.05 \leq \alpha \leq 0.14$, Methods and Supplementary Fig. 1). Using this design we identified 501 putative genetic interactions influencing the expression levels of 238 genes (Supplementary Table 1). We used strict quality control measures to avoid statistical associations being driven by technical artefacts (Methods). However, it remains possible that unexplained technical artefacts may have led to the significant discovery interactions. Of the 501 discovery interactions, 434 had available data and passed filtering (Methods) in two independent replication data sets, Fehrmann[12] and the Estonian Genomics Centre University of Tartu (EGCUT)[11], in which we saw convincing evidence for replication. We used the summary statistics from the replication data sets to perform a meta-analysis to obtain an independent P value for the putative interactions, and 30 were significant after applying a Bonferroni correction for multiple testing (5% significance threshold $P < 9.98 \times 10^{-5}$, Table 1). To quantify the similarity of genotype–phenotype maps between the independent data sets (Fig. 1) we decomposed the genetic effects of each of the SNP pairs into orthogonal additive, dominance and epistatic effects (A1, A2, D1, D2, A1 × A2, A1 × D2, D1 × A2, D1 × D2) and tested for concordance of the sign of the most statistically significant effect (Supplementary

[1]Queensland Brain Institute, University of Queensland, Brisbane, Queensland 4072, Australia. [2]University of Queensland Diamantina Institute, University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland 4072, Australia. [3]Department of Genetics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9700 RB Groningen, The Netherlands. [4]Estonian Genome Center, University of Tartu, Tartu 51010, Estonia. [5]Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA. [6]Divisions of Endocrinology, Children's Hospital, Boston, Massachusetts 02115, USA. [7]Queensland Institute of Medical Research, Brisbane, Queensland 4006, Australia. [8]School of Biology and Centre for Integrative Genomics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.
*These authors contributed equally to this work.

Table 1 | Epistatic interactions significant at the Bonferroni level in two replication sets

	Gene (chr.)	SNP 1 (chr.)	SNP 2 (chr.)	BSGS†	Fehrmann‡	EGCUT‡	Meta§
1	ADK (10)	rs2395095 (10)	rs10824092 (10)	6.69*	18.33*	21.21*	39.82*
2	ATP13A1 (19)	rs4284750 (19)	rs873870 (19)	5.30	12.18	3.25	14.23
3	C21ORF57 (21)	rs9978658 (21)	rs11701361 (21)	9.42	6.08	16.36	21.67
4	CSTB (21)	rs9979356 (21)	rs3761385 (21)	11.99	25.20	16.72	42.27
5	CTSC (11)	rs7930237 (11)	rs556895 (11)	7.16	18.76	15.06	33.53
6	FN3KRP (17)	rs898095 (17)	rs9892064 (17)	16.16	28.24	29.39	59.95
7	GAA (17)	rs11150847 (17)	rs12602462 (17)	13.91	19.98	12.99	32.60
8	HNRPH1 (5)	rs6894268 (5)	rs4700810 (5)	15.38	8.55	3.01	10.37
9	LAX1 (1)	rs1891432 (1)	rs10900520 (1)	19.16	18.60	11.22	29.24
10	MBNL1 (3)	rs16864367 (3)	rs13079208 (3)	13.49	16.25	24.74	41.56
11	MBNL1 (3)	rs7710738 (5)	rs13069559 (3)	7.92	2.55	7.89	9.28
12	MBNL1 (3)	rs2030926 (6)	rs13069559 (3)	7.10	0.91	5.80	5.53
13	MBNL1 (3)	rs2614467 (14)	rs13069559 (3)	5.74	4.13	2.22	5.30
14	MBNL1 (3)	rs218671 (17)	rs13069559 (3)	7.63	0.62	5.82	5.23
15	MBNL1 (3)	rs11981513 (7)	rs13069559 (3)	7.71	0.43	5.36	4.58
16	MBP (18)	rs8092433 (18)	rs4890876 (18)	5.40	7.06	21.91	28.73
17	NAPRT1 (8)	rs2123758 (8)	rs3889129 (8)	8.45	15.12	16.08	30.77
18	NCL (2)	rs7563453 (2)	rs4973397 (2)	7.31	7.51	6.33	12.70
19	PRMT2 (21)	rs2839372 (21)	rs11701058 (21)	4.81	0.69	4.47	4.06
20	RPL13 (16)	rs352935 (16)	rs2965817 (16)	4.98	3.79	14.41	17.24
21	SNORD14A (11)	rs2634462 (11)	rs6486334 (11)	7.31	13.11	10.96	23.22
22	TMEM149 (19)	rs807491 (19)	rs7254601 (19)	12.16	81.55	45.78	145.78
23	TMEM149 (19)	rs8106959 (19)	rs6926382 (6)	5.80	3.06	8.80	10.72
24	TMEM149 (19)	rs8106959 (19)	rs914940 (1)	6.22	3.36	6.96	9.20
25	TMEM149 (19)	rs8106959 (19)	rs2351458 (4)	7.30	0.04	9.61	8.00
26	TMEM149 (19)	rs8106959 (19)	rs6718480 (2)	8.55	3.31	5.15	7.36
27	TMEM149 (19)	rs8106959 (19)	rs1843357 (8)	6.21	3.72	3.33	6.00
28	TMEM149 (19)	rs8106959 (19)	rs9509428 (13)	9.44	0.10	5.75	4.47
29	TRA2A (7)	rs7776572 (7)	rs11770192 (7)	8.23	3.19	1.89	4.09
30	VASP (19)	rs1264226 (19)	rs2276470 (19)	5.09	0.94	5.14	4.95

* $-\log_{10} P$ values for 4 d.f. interaction tests
† Discovery data set
‡ Independent replication data set
§ Meta-analysis of interaction terms between replication data sets only

Table 3 and Methods). Sign concordance between the discovery and both replication data sets was observed in 22 out of the 30 significantly replicated interactions (expected value = 7.5 under the null hypothesis of no interactions, $P = 3.76 \times 10^{-8}$).

In addition, using the meta-analysis from the replication samples only, we observed that 316 of the remaining 404 discovery SNP pairs had replication interaction P values more extreme than the 2.5% confidence interval of the quantile–quantile plot against the null hypothesis of no interactions in which P values are assumed to be uniformly distributed ($P << 1.0 \times 10^{-16}$, Fig. 2 and Supplementary Fig. 2). Concordance of the direction of the effect of the largest variance component was also highly significant ($P = 5.71 \times 10^{-31}$, Supplementary Table 3). The congruence of the epistatic networks in discovery and replication data sets is shown in Fig. 3, demonstrating that these complex genetic patterns are common even across independent data sets. A further replication was attempted using the Centre for Health Discovery and Wellbeing (CHDWB) data set[24], but only 20 of the SNP pairs passed filtering because the sample size was small ($n = 139$), and probably because of insufficient power we found no evidence for replication (Supplementary Fig. 6). It should be noted that although it is a necessary step to establish the veracity of the interactions from the discovery set, replication of epistatic effects in independent samples is difficult in practice due to linkage disequilibrium (Methods).

Although seldom the focus of association studies, SNPs with known main effects are often tested for $A \times A$ genetic interactions[9], but our analysis suggests that this is unlikely to be the best strategy for its detection. The majority of our discovery interactions were composed of one SNP that was significantly associated with the gene expression level in the discovery data set, and one SNP that had no previous association[22] (439 out of 501, Methods). Only nine interactions were between SNPs that both had known main effects, whereas 64 were between SNPs that had no known main effects. Additionally, we observed that the largest epistatic variance component for the 501 interactions was equally divided amongst $A \times A$, $A \times D$, $D \times A$ and $D \times D$ at the

discovery stage ($P = 0.22$ for departure from expectation). This is not surprising because these patterns of epistasis used for statistical decomposition are simply convenient orthogonal parameterizations of a two-locus model, and are not intended to model biological function[25].

Of the discovery interactions, 26 were cis–cis acting (within 1 megabase (Mb) of the transcription start site, mean distance between SNPs was 0.53 Mb), 462 were cis–trans-acting and 13 were trans–trans-acting. We observed a wide range of significant genotype–phenotype maps (Fig. 1) but the most common pattern of epistasis that we detected involved a trans-SNP masking the effect of an additive cis-SNP. For example, MBNL1 (involved in RNA modification and regulation of splicing[26]) has a cis effect at rs13069559 which in turn is controlled by 13 trans-SNPs and one cis-SNP that each exhibit a masking pattern, such that when the trans-SNP is homozygous for the masking allele the decreasing allele of the cis-SNP no longer has an effect (Supplementary Fig. 10). Each of these interactions has evidence for replication in at least one data set and six are significantly replicated at the Bonferroni level (Supplementary Fig. 3). We see similar epistatic networks involving multiple (eight or more) trans-acting SNPs for other gene expresson levels too, for example TMEM149 (also known as IGFLR1) (Supplementary Fig. 11), NAPRT1 (Supplementary Fig. 12), TRAPPC5 (Supplementary Fig. 13) and CAST (Supplementary Fig. 14). We observed from pedigree analysis that these five gene expression phenotypes had non-additive variance component estimates within the 95th percentile of the 17,994 gene expression phenotypes that were analysed previously[22] (Supplementary Table 2 and Methods).

In total the 501 interactions comprised 781 unique SNPs, which we analysed for functional enrichment (Methods). We tested the SNPs for cell-type specific overlap with transcriptionally active chromatin regions, tagged by histone-3-lysine-4, tri-methylation (H3K4me3) chromatin marks, in 34 cell types[27] (Supplementary Fig. 5). There was significant enrichment for cis-acting SNPs in haematopoietic cell types only ($P < 1 \times 10^{-4}$ for the three tissues with the strongest enrichment after adjusting for multiple testing). However trans-acting SNPs did not

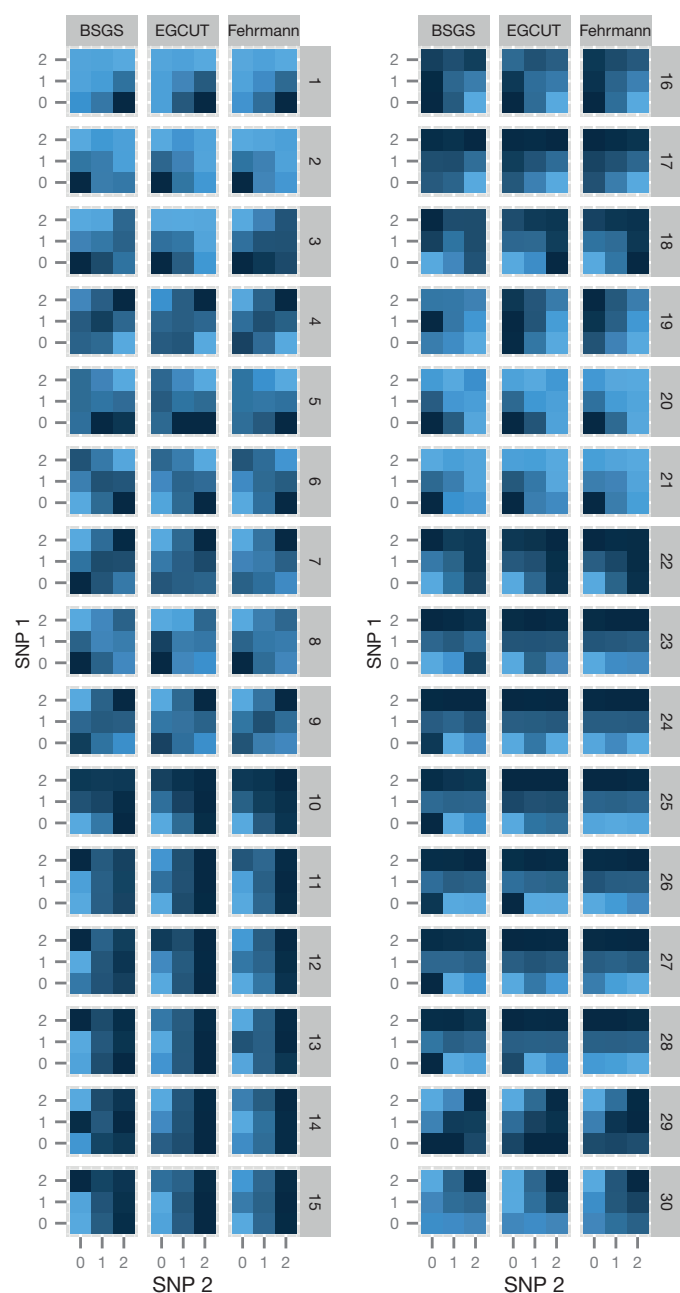**Figure 1 | Replication of genotype–phenotype maps in two independent populations.** The genotype–phenotype maps for each epistatic interaction that is significant at the Bonferroni level in both replication data sets are shown. Each genotype–phenotype map consists of nine tiles where each tile represents the expression level for that two-locus genotype class. Phenotypes are for gene transcript levels (dark coloured tiles = high expression, light coloured tiles = low expression). Columns of genotype–phenotype maps are for each independent data set. Rows of genotype–phenotype maps are for each of 30 significantly replicated interactions at the Bonferroni level, corresponding to the rows in Table 1. There is a clear trend of the genotype–phenotype maps replicating across all three data sets.

Figure 2 | Q-Q plots of interaction P values from replication data sets. The top panel shows all 434 discovery SNPs that were tested for interactions. Observed P values (y axis, −log$_{10}$ scale) are plotted against the expected P values (x axis, −log$_{10}$ scale). The multiple testing correction threshold for significance following Bonferroni correction is denoted by a dotted line. The bottom panel shows the same data as the top panel but excluding the 30 interactions that were significant at the Bonferroni level in the replication data sets. The shaded grey area represents the 5% confidence interval for the expected distribution of P values. Dark blue points represent P values that exceed the confidence interval, light blue are within the confidence interval.

show any tissue specific enrichment ($P > 0.1$ for all tissues). This difference between *cis* and *trans* SNPs suggests different roles in epistatic interactions where tissue specificity is provided by the *cis*-SNPs. There is also enrichment for *cis*-SNPs to be localized in regions with regulatory genomic features as measured by chromatin states[28] (Supplementary Fig. 4).

We also demonstrate physical organization of interacting loci within the cell, suggesting a mechanism by which biological function can lead to

epistatic genetic variance. It has been shown that different chromosomal regions spatially co-localize in the cell through chromatin interactions[13]. We cross-referenced our epistatic SNPs with a map of chromosome interacting regions ($n = 96,139$) in K562 blood cell lines[29] (Methods) and found that 44 epistatic interactions mapped to within 5 Mb ($P < 1.8 \times 10^{-10}$),
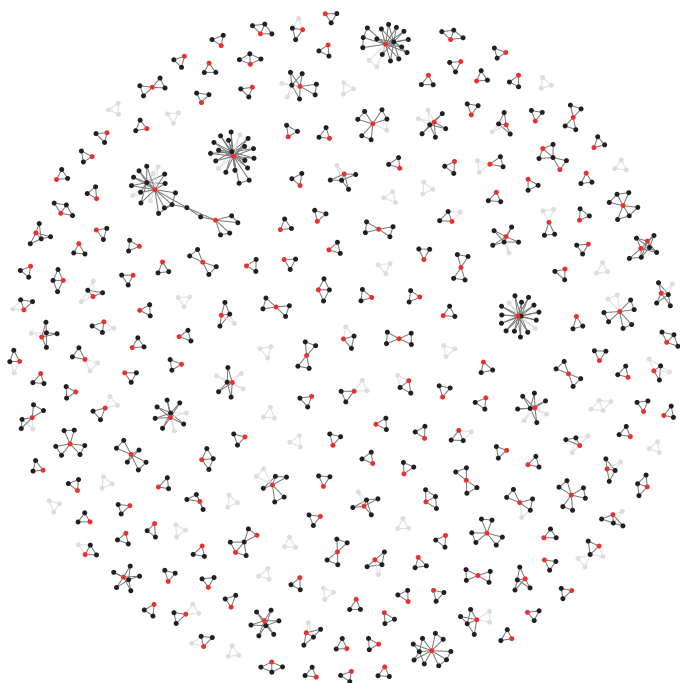
Figure 3 | Discovery and replication of epistatic networks. All 434 putative genetic interactions (edges) with data common to discovery and replication sets is shown, where black nodes represent SNPs and red nodes represent traits (gene expression probes). Three hundred and forty-five interactions had *P* values exceeding the 2.5% confidence interval following meta-analysis of the replication data. The remaining 89 interactions that did not replicate are depicted in grey. It is evident that a large proportion of the complex networks identified in the discovery set also exist in independent populations. An interactive version of this graph can be found at (http://kn3in.github.io/detecting_epi/).

(Supplementary Fig. 15). Interaction of distant loci may occur through physical proximity in transcriptional factories that organize across different chromosome regions and can regulate transcription of related genes[30].

Quantifying the importance of epistasis in complex traits in humans remains an open question. Here we are able to identify 238 gene expression traits with at least one significant interaction given our experiment-wide threshold, where the minimum estimated variance explained by the epistatic effects of any interaction was 2.1% of phenotypic variance. Taking results from our previously published expression quantitative trait loci (eQTL)[23], we calculated that 1,848 of the 7,339 gene expression levels analysed were influenced by additive effects where the estimated additive variance of a locus was 2.1% or greater. Thus, we can infer that the number of instances of large additive effects is substantially greater than the number of instances of large epistatic effects.

In terms of their contribution to complex traits, a more important metric might be the proportion of the variance that the epistatic loci explain[2]. Taking all additive effects detected in ref. 23 that have additive variance explaining 2.1% or greater of phenotypic variance, we calculated that the proportion of total phenotypic variance of all 7,339 gene expression levels explained by additive effects alone was 2.16%. By contrast, the estimated epistatic variance from the interacting SNPs detected in this study on average explain a total of 0.22% of phenotypic variance, approximately ten times lower than the estimated additive variance. There are several caveats to this comparison which we discuss in the Methods.

Overall, we have demonstrated that it is possible to identify and replicate epistasis in complex traits amongst common human variants, despite the relative contribution of pairwise epistasis to phenotypic variation being small. The bioinformatic analyses of the significant epistatic loci indicate that there are a large number of possible mechanisms that can lead to non-additive genetic variation. Further research into such

epistatic effects may provide a useful framework for understanding molecular mechanisms and complex trait variation in greater detail. With computational techniques and data now widely available, the search for epistasis in larger data sets for traits of broader interest is warranted.

METHODS SUMMARY

We searched for pairwise epistasis exhaustively in the BSGS discovery data set[23], which comprises 846 individuals who are genotyped at 528,509 autosomal SNPs. Each individual had gene expression levels measured in peripheral blood at 7,339 probes representing 6,158 RefSeq genes (significant expression in $\geq$ 90% of individuals). SNP pairs were modelled for full genetic effects, including marginal additive and dominance at both SNPs plus four interaction terms. We used permutation analysis to calculate an experiment-wide significance threshold of $T_e = 2.91 \times 10^{-16}$ at the 5% family-wise error rate (FWER). All SNP pairs with linkage disequilibrium $r^2 > 0.1$ and $D'^2 > 0.1$ were removed to minimise the possibility of haplotype effects. All SNP pairs were required to have at least five data points in all nine genotype classes. If multiple SNP pairs were present on the same chromosomes for a particular expression trait then only the sentinel SNP pair was retained. Finally, a nested test contrasting the full genetic model against the marginal additive and dominance model was performed for each remaining SNP pair. The 501 significant SNP pairs were carried forward for replication in two independent data sets that used the same expression assays for analysing transcription in peripheral blood Fehrmann[12] ($n = 1,240$) and EGCUT[11] ($n = 891$). A meta-analysis on the interaction *P* values from each replication data set was performed to provide an overall replication statistic for each putative interaction.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Carlborg, O. & Haley, C. S. Epistasis: too often neglected in complex trait studies? *Nature Genetics* **5**, 618–625 (2004).
2. Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008 (2008).
3. Crow, J. F. On epistasis: why it is unimportant in polygenic directional selection. *Phil. Trans. R. Soc. B* **365**, 1241–1244 (2010).
4. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
5. Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* 234–237 (2013).
6. Carlborg, O., Jacobsson, L., Ahgren, P., Siegel, P. & Andersson, L. Epistasis and the release of genetic variation during long-term selection. *Nature Genetics* **38**, 418–420 (2006).
7. Strange, A. *et al.* A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAP1*. *Nature Genetics* **42**, 985–990 (2010).
8. Evans, D. M. *et al.* Interaction between *ERAP1* and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nature Genetics* **43**, (2011).
9. Cordell, H. J. Detecting gene–gene interactions that underlie human diseases. *Nature Rev. Genetics* **10**, 392–404 (2009).
10. Hemani, G., Theocharidis, A., Wei, W. & Haley, C. EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics* **27**, 1462–1465 (2011).
11. Metspalu, A. The Estonian Genome Project. *Drug Dev. Res.* **62**, 97–101 (2004).
12. Fehrmann, R. S. N. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genetics* **7**, e1002197 (2011).
13. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
14. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
15. Weinreich, D. M., Delaney, N. F., Depristo, M., a. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
16. Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. a. Epistasis as the primary factor in molecular evolution. *Nature* **490**, 535–538 (2012).
17. Weir, B. S. Linkage disequilibrium and association mapping. *Annu. Rev. Genomics Hum. Genet.* **9**, 129–142 (2008).
18. Hemani, G., Knott, S. & Haley, C. An evolutionary perspective on epistasis and the missing heritability. *PLoS Genet.* **9**, e1003295 (2013).
19. Marchini, J., Donnelly, P. & Cardon, L. R. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature Genet.* **37**, 413–417 (2005).
20. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
21. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).

22. Powell, J. E. *et al.* Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. *PLoS Genet.* **9,** e1003502 (2013).
23. Powell, J. E. *et al.* The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS ONE* **7,** e35430 (2012).
24. Preininger, M. *et al.* Blood-informative transcripts define nine common axes of peripheral blood gene expression. *PLoS Genet.* **9,** e1003362 (2013).
25. Cockerham, C. C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39,** 859–882 (1954).
26. Ho, T. H. *et al.* Muscleblind proteins regulate alternative splicing. *EMBO J.* **23,** 3103–3112 (2004).
27. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature Genet.* **45,** 124–130 (2013).
28. Hoffman, M. M., Buske, O., Wang, J. & Weng, Z. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nature Methods* **9,** 473–476 (2012).
29. Lan, X. *et al.* Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Res.* **40,** 7690–7704 (2012).
30. Rieder, D., Trajanoski, Z. & McNally, J. G. Transcription factories. *Front. Genet.* **3,** 221 (2012).

METHODS

Discovery data. The Brisbane Systems Genetics Study (BSGS) comprises 846 individuals of European descent from 274 independent families[23]. DNA samples from each individual were genotyped on the Illumina 610-Quad Beadchip by the Scientific Services Division at deCODE Genetics Iceland. Full details of genotyping procedures are given in ref. 31. Standard quality control (QC) filters were applied and the remaining 528,509 autosomal SNPs were carried forward for further analysis.

Gene expression profiles were generated from peripheral blood collected with PAXgene TM tubes (Qiagen, Valencia, CA) using Illumina HT12-v4.0 bead arrays. The Illumina HT-12 v4.0 chip contains 47,323 probes, although some probes are not assigned to RefSeq genes. We removed any probes that did not match the following criteria: contained a SNP within the probe sequence with MAF >0.05 within 1,000 genomes data; did not map to a listed RefSeq gene; were not significantly expressed (based on a detection P value <0.05) in at least 90% of samples. After this stringent QC, 7,339 probes remained for 2D-eQTL mapping. These data are accessible through GEO accession number GSE53195.

Normalization: gene expression profiles were normalized and adjusted for batch and polygenic effects. Profiles were first adjusted for raw background expression in each sample. Expression levels were then adjusted using quantile and $\log_2$ transformation to standardise distributions between samples. Batch and polygenic effects were adjusted using the linear model:

$$y = \mu + \beta_1 c + \beta_2 p + \beta_3 s + \beta_4 \alpha + g + e \tag{1}$$

where μ is the population mean expression levels, c, p, s and a are vectors of chip, chip position, sex and generation respectively, fitted as fixed effects; and g is a random additive polygenic effect with a variance covariance matrix

$$G_{jk} = \begin{cases} \sigma_a^2 & j = k \\ 2\phi_{jk}\sigma_a^2 & j \neq k \end{cases} \tag{2}$$

The parameter σ_a^2 is the variance component for additive background genetic. Here we are using family based pedigree information rather than SNP based IBD to account for relationships between individuals and so ϕ_{jk} is the kinship coefficient between individuals j and k. The residual, e, from equation (1) is assumed to follow a multivariate normal distribution with a mean of zero. Residuals were normalized by rank transformation and used as the adjusted phenotype for the pairwise epistasis scan to remove any skewness and avoid results being driven by outliers. The GenABEL package for R was used to perform the normalization[32].

Exhaustive 2D-eQTL analysis. Two-stage search: we used epiGPU[10] software to perform an exhaustive scan for pairwise interactions, such that each SNP is tested against all other SNPs for statistical association with the expression values for each of the 7,339 probes. This uses the massively parallel computational architecture of graphical processing units (GPUs) to speed up the exhaustive search. For each SNP pair there are 9 possible genotype classes. We treat each genotype class as a fixed effect and fit an 8 d.f. F-test to test the following hypotheses:

$$H_0 : \sum_{i=1}^{3}\sum_{j=1}^{3} \left(\bar{x}_{ij} - \mu\right)^2 = 0; \tag{3}$$

$$H_1 : \sum_{i=1}^{3}\sum_{j=1}^{3} \left(\bar{x}_{ij} - \mu\right)^2 = 0; \tag{4}$$

where μ is the mean expression level and x_{ij} is the pairwise genotype class mean for genotype i at SNP 1 and genotype j at SNP 2. This type of test does not parameterize for specific types of epistasis, rather it tests for the joint genetic effects at two loci. This has been demonstrated to be statistically more efficient when searching for a wide range of epistatic patterns, although will also include any marginal effects of SNPs which must be dealt with post-hoc[18].

Stage 1: the complete exhaustive scan for 7,339 probes comprises 1.03×10^{15} F-tests. We used permutation analysis to estimate an appropriate significance threshold for the study. To do this we performed a further 1,600 exhaustive 2D scans on permuted phenotypes to generate a null distribution of the extreme P values expected to be obtained from this number of multiple tests given the correlation structure between the SNPs. We took the most extreme P value from each of the 1,600 scans and set the 5% family-wise error rate (FWER) to be the 95% most extreme of these P values, $T_* = 2.13 \times 10^{-12}$. The effective number of tests in one 2D scan being performed is therefore $N_* = 0.05/T_* \approx 2.33 \times 10^{10}$. To correct for the testing of multiple traits we established an experiment wide threshold of $T_e = 0.05/(N_* \times 7339) = 2.91 \times 10^{-16}$. This is likely to be conservative as it assumes independence between probes.

Filtering: we used two approaches to filter SNPs from stage 1 to be tested for significant interaction effects in stage 2.

Filter 1: after keeping SNP pairs that surpassed the 2.91×10^{-16} threshold in stage 1 only SNP pairs with at least 5 data points in all 9 genotype classes were kept. We then calculated the linkage disequilibrium between interacting SNPs (amongst unrelated individuals within the discovery sample and also from 1,000 genomes data) and removed any pairs with $r^2 > 0.1$ or $D'^2 > 0.1$ to avoid the inclusion of haplotype effects and to increase the accuracy of genetic variance decomposition. If multiple SNP pairs were present on the same chromosomes for a particular expression trait then only the sentinel SNP pair was retained, that is, if a probe had multiple SNP pairs that were on chromosomes one and two then only the SNP pair with the most significant P value was retained. At this stage, 6,404 filtered SNP pairs remained.

Filter 2: we also performed a second filtering screen applied to the list of SNP pairs from stage 1 that was identical to filter 1 but an additional step was included in which any SNPs that had previously been shown to have a significant additive or dominant effect ($P < 1.29 \times 10^{-11}$) were removed[22], creating a second set of 4,751 unique filtered SNP pairs.

Stage 2: to ensure that interacting SNPs were driven by epistasis and not marginal effects, we performed a nested ANOVA on each pair in the filtered set to test if the interaction terms were significant. We did this by contrasting the full genetic model (8 d.f.) against the reduced marginal effects model which included the additive and dominance terms at both SNPs (4 d.f.). Thus, a 4 d.f. F-test was performed on the residual genetic variation, representing the contribution of epistatic variance. Significance of epistasis was determined using a Bonferroni threshold of $0.05/(6404 + 4751) = 4.48 \times 10^{-6}$. This resulted in 406 and 95 SNP pairs with significant interaction terms from filters 1 and 2, respectively.

Type 1 error rate. Using a Bonferroni correction of 0.05 in the second stage of the two stage discovery scan implies a type 1 error rate of $\alpha = 0.05$. However, this could be underestimated because the number tests performed in the second stage depends on the number of tests in the first stage, and this depends on statistical power and model choice. We performed simulations to estimate the type 1 error rate of this study design.

We assumed a null model where there was one true additive effect and 7 other terms with no effect. To simulate a test statistic we simulated 8 z-scores, $z_1 \sim N(\sqrt{NCP}, 1)$ and $z_{2..8} \sim N(0, 1)$. Thus $z_{full} = \sum_{i=1}^{8} z_i \sim \chi_8^2$ (representing the 8 d.f. test) and $z_{int} = \sum_{i=5}^{8} z_i \sim \chi_4^2$ (representing the 4 d.f. test in which the null hypothesis of no epistasis is true). For a particular value of NCP we simulated 100,000 z values, and calculated the P_{full} value for the z_{full} test statistic. The n_{int} test statistics with $p_{full} < 2.31 \times 10^{-16}$ were kept for the second stage, where the type 1 error rate of stage 2 was calculated as the proportion of $P_{int} < 0.05/n_{int}$. The power at stage 1 was calculated as $n_{int}/100,000$. This procedure was performed for a range of NCP parameters that represented power ranging from ~ 0 to ~ 1.

Population stratification. We ruled out population stratification as a possible cause of inflated test statistics. To test for cryptic relatedness driving the interaction terms we tested for increased linkage disequilibrium among the SNPs[33]. We calculated the mean of the off-diagonal elements of the correlation matrix of all unique SNPs from the 501 interactions (731 SNPs) using only unrelated individuals, $\overline{r^2} = 0.0039$. This is not significantly different from the null hypothesis of zero (sampling error $= 1/n_{unrelated} = 0.0039$).

Probe mapping. To avoid possibility that epistatic signals might arise due to expression probes hybridizing in multiple locations we verified that probe sequences for genes with significant interactions mapped to only a single location. As an initial verification we performed a BLAST search of the full probe sequence against 1,000 genomes phase 1 version 3 human genome reference and ensured that only one genomic location aligned significantly ($P < 0.05$). As a second step, to mitigate the possibility of weak hybridization elsewhere in the genome we divided the probe sequence into three sections (1–25 bp, 13–37 bp, 26–50 bp) and performed a BLAST search of these probe sequence fragments. No probe sequences or probe sequence fragments mapped to positions other than the single expected genomic target ($P < 0.05$).

Replication. Data description: we attempted replication of the 501 significant interactions from the discovery set using three independent cohorts; Fehrmann, EGCUT, and CHDWB. It was required that linkage disequilibrium $r^2 < 0.1$ and $D'^2 < 0.1$ between interacting SNPs (as measured in the replication sample directly), and all nine genotype classes had at least 5 individuals present in order to proceed with statistical testing for replication in both data sets. We also excluded any putative SNPs that had discordant allele frequencies in any of the data sets. Details of the cohorts are as follows.

Fehrmann: the Fehrmann data set[12] consists of peripheral blood samples of $n = 1,240$ unrelated individuals from the UK and the Netherlands. Some of these individuals are patients and others are healthy controls.

Individuals were genotyped using the Illumina HumanHap300, Illumina Human-Hap370CNV, and Illumina 610 Quad platforms. RNA levels were quantified using

the Illumina HT-12 V3.0 platform. These data are accessible through GEO accession numbers GSE20332 and GSE20142.

EGCUT: the Estonian Genome Center of the University of Tartu (EGCUT) study[11] consists of peripheral blood samples of $n = 891$ unrelated individuals from Estonia. They were genotyped using the Illumina HumanHap370CNV platform. RNA levels were quantified using the Illumina HT-12 V3.0 platform. These data are accessible through GEO accession number GSE48348.

CDHWB: the Center for Health Discovery and Well Being (CDHWB) Study[24] is a population based cohort consisting of $n = 139$ individuals of European descent collected in Atlanta, USA. Gene expression profiles were generated with Illumina HT-12 V3.0 arrays from peripheral blood collected from Tempus tubes that preserve RNA. Whole genome genotypes were measured using Illumina OmniQuad arrays. Owing to the small sample size, most SNP pairs did not pass filtering in this data set (20 SNP pairs remained) and so we have excluded it from the rest of the analysis.

Meta-analysis. The 4 d.f. interaction P values for each independent replication data set were calculated using the same statistical test as was performed in the discovery data set. We then took the interaction P values from EGCUT and Fehrmann and calculated a joint P value using Fisher's method of combining P values for a meta-analysis as $-2 \ln p_1 - 2 \ln p_2 \sim \mathcal{X}_{4\mathrm{d.f.}}^2$. As in the discovery analysis, all gene expression levels were normalized using rank transformation to avoid skew or outliers in the distribution[34].

Concordance of direction of effects. We used four methods to calculate the concordance of the direction of effects between the discovery and replication data sets.

Test 1: here we test whether the most significant epistatic effect in the discovery set is in the same direction as the same epistatic effect in the replication sets. We decomposed the genetic variance into 8 orthogonal effects, four of which are epistatic ($A \times A$, $A \times D$, $D \times A$, $D \times D$). The sign of the epistatic effect that had the largest variance in the discovery was recorded, and then was compared to the same epistatic effect in the two replication data sets (regardless of whether or not the same epistatic effect was the largest in the replication data sets). The probability of the sign being the same in one data set is 1/2. The probability of the sign being the same in two is 1/4.

Test 2: here we test whether the most significant epistatic effect in the discovery is the same as the largest epistatic effect in the replication set with the sign being concordant. As in Test 1, but this time we required that the largest effect was the same in the discovery and the replication, and that they had the same sign (for example, if the largest effect in the discovery is $A \times A$, with a positive effect, then concordance is achieved if the same is true in the replication). The probability of one replication data set being concordant by chance is 1/8, and concordance in both is 1/64.

Test 3: here we test if the epistatic effects that are significant at nominal $P < 0.05$ in the discovery have the same direction of effect as in the replication. We count all the epistatic variance components in the discovery that have $P < 0.05$ (1,133 amongst the 434 discovery SNP pairs, that is, each SNP pair has at least 1 and at most 4 significant epistatic variance components). Then we compare the direction of the effect in the replication data set. The probability of the sign being the same in one data set for any one significant effect is 1/2. The probability of the sign being the same in two is 1/4.

Test 4: if we count how many of the 4 epistatic effects are concordant between the discovery and replication data for each interaction then is this significant from what we expect by chance? There can be either 0, 1, 2, 3 or 4 concordant signs at each interaction, each with expectation of $P = 1/16, 4/16, 6/16, 4/16, 1/16$ under the null hypothesis, respectively. Observed counts are multinomially distributed, and we tested whether the observed proportions were statistically different from the expected proportions using an approximation of the multinomial test[35].

The probability of observing the number of concordant signs in tests 1–3 is calculated using a binomial test. All variance decompositions were calculated using the NOIA method[36].

Effects of linkage disequilibrium on detection and replication. The power to detect genetic effects, when the observed markers are in linkage disequilibrium with the causal variants, is proportional to r^x. For additive effects $x = 2$, but for non-additive effects x is larger, that is, $x = 4$ for dominance or $A \times A$, $x = 6$ for $A \times D$ or $D \times A$, and $x = 8$ for $D \times D$. Many biologically realistic genotype–phenotype maps may be comprised of all 8 variance components[18].

This is important for both detection and for replication of epistasis. For detection, if the epistatic effect includes the $D \times D$ term then if the two causal variants are tagged by observed markers that are each in linkage disequilibrium $r = 0.9$, then if the true variance is V_t then the observed variance V_o at the markers will be $0.9^8 V_t = 0.43 V_t$. Therefore, it is important to consider the sampling variation of $\hat{r}^x$ in a sample given some true population value of r.

Simulation 1: for some values of fixed population parameters, p_1 (minor allele frequency at observed marker), q_1 (minor allele frequency at causal variant), and r

(linkage disequilibrium between marker and causal variant), the expected haplotype frequencies are:

$$h_{11} = r\sqrt{p_1 q_1 p_2 q_2} + p_1 q_1 \tag{5}$$

$$h_{12} = p_1 q_2 - r\sqrt{p_1 q_1 p_2 q_2} \tag{6}$$

$$h_{21} = p_2 q_1 - r\sqrt{p_1 q_1 p_2 q_2} \tag{7}$$

$$h_{22} = r\sqrt{p_1 q_1 p_2 q_2} + p_2 q_2 \tag{8}$$

where $p_2 = 1 - p_1$ and $q_2 = 1 - q_1$. For a range of population parameters we randomly sampled $2n$ haplotypes where the expected haplotype frequencies were $h_{11}, h_{12}, h_{21}, h_{22}$. From the sample haplotype frequencies we then calculated sample estimates of $\hat{r}$ where:

$$\hat{r} = \frac{\hat{h}_{11} - \hat{p}_1 \hat{q}_1}{\sqrt{\hat{p}_1 \hat{q}_1 \hat{p}_2 \hat{q}_2}} \tag{9}$$

For each value of combination of the parameters p_1, q_1, r and n, 1,000 simulations were performed and the sampling mean and sampling standard deviation of $\hat{r}, \hat{r}^2, \hat{r}^4, \hat{r}^6, \hat{r}^8$ were recorded. It was observed that sampling variance increases for increasing x in $\hat{r}^x$.

Simulation 2: we assume that the discovery SNP pairs are ascertained (from a very large number of tests) to have high $\hat{r}$ between observed SNPs and causal variants because otherwise power of detection would be low. We can speculate that the distribution of $\hat{r}$ in this ascertained sample will be a mixture of r that is high and r that is lower but with ascertained higher values from sampling. Therefore, we would expect those with truly high r to have a higher replication rate in independent data sets, and those with ascertained high $\hat{r}$ to have lower replication because resampling is unlikely to result in the same extreme ascertainment. To obtain empirical estimates of $\hat{r}$ in discovery and replication data sets we conducted the following simulation: using 1,000 genomes data (phase 1, version 3, 379 European samples) we selected the 528,509 'markers' used in the original discovery analysis, plus 100,000 randomly chosen 'causal variants' (CVs) with minor allele frequency > 0.05; the 379 individuals were split into discovery (190) and replication (189) sets; for each CV, the marker with the maximum $\hat{r}_D^2$ from the marker panel was recorded in the discovery set (this marker was known as the 'discovery marker' (DM)); the $\hat{r}_R^2$ for each CV/DM pair was then calculated in the replication set where the discovery linkage disequilibrium was ascertained to be high, such that $\hat{r}_D^2 > 0.9$.

We observed that there was an average decrease in $\hat{r}_R^x$ relative to $\hat{r}_D^x$, and that this decrease was larger with increasing x. We observed that $\left(\hat{r}_R^2 - \hat{r}_D^2\right)/\hat{r}_D^2 = 0.029$ Whereas $\left(\hat{r}_R^8 - \hat{r}_D^8\right)/\hat{r}_D^8 = 0.092$. The average drop in in replication $\hat{r}^8$ was 3 times higher than the drop in $\hat{r}^2$.

Interpretation: simulation 1 shows that sampling variance of r^x increases as x increases. Detection of epistatis is highly dependent upon high $\hat{r}$. Amongst the discovery SNPs, there will be a mixture of interactions where observed SNPs are either in true high linkage disequilibrium with causal variants or will have highly inflated sample $\hat{r}^x$ compared to the population r^x. Simulation 2 shows that as x gets larger, the average decrease in $\hat{r}^x$ between discovery and replication becomes larger, probably a result of ascertained high $\hat{r}$ in the discovery and increased sampling variance with increasing x in the replication. These results demonstrate that if all else is equal, the impact of sampling variance of r alone will reduce the replication rate of epistatic effects compared to additive effects.

Additive and non-additive variance estimation. Fixed effects: to compare the relative contribution to the phenotypic variance of gene expression levels between additive and epistatic effects we are constrained by the problem that non-additive variance components for a phenotype cannot be calculated directly. Here, we only have SNP pairs that exceed a threshold of $P < 2.91 \times 10^{-16} = T_e$. A strong conclusion cannot be made about the genome-wide variance contribution, but we can compare the variance explained by SNP effects at this threshold for additive scans and epistatic scans.

In ref. 23, an expression quantitative trait locus (eQTL) study was performed searching for additive effects in the same BSGS data set as was used for the discovery here. Using the threshold T_e for the additive eQTL study, 453 of the 7,339 probes analysed here had at least one significant additive effect. Assuming that the phenotypic variance for each of the probes is normalized to 1, the total phenotypic variance of all 7,339 explained by the significant additive effects was 1.73%.

Following the same procedure, at the threshold T_e there were 238 gene expression probes with at least one significant pairwise epistatic interaction out of the 7,339 tested. In total the proportion of the phenotypic variance explained by the epistatic effects at these SNP pairs was 0.25%.

Limitations of this type of comparison: although it is useful to compare the relative variances of epistatic and additive effects, it must be stressed that our

results here are approximations that are very limited by the study design. We estimate that additive effects explain approximately 10 times more variance than epistatic effects, but this could be an overestimate or an underestimate owing to a number of different caveats. First, the ratio of additive to epistatic variance may differ at different minimum variance thresholds, and our estimate is determined by the threshold used. Second, the power of a 1 d.f. test exceeds that of an 8 d.f. test. Third, the non-additive variance at causal variants is expected to be underestimated by observed SNPs in comparison to estimates for additive variance. Last, the extent of winner's curse in estimation of effect sizes may differ between the two studies.

Pedigree estimates. The gene expression levels for *MBNL1*, *TMEM149*, *NAPRT1*, *TRAPPC5* and *CAST* are influenced by large *cis–trans* epistatic networks (eight interactions or more). Though it is not possible to orthogonally estimate the non-additive genetic variance for non-clonal populations, an approximation of a component of non-additive variance can be estimated using pedigree information. The BSGS data are comprised of some related individuals and standard quantitative genetic analysis was used to calculate the additive and dominance variance components for each gene expression phenotype in ref. 22. The dominance effect is likely to capture additive × additive genetic variance plus some fraction of other epistatic variance components. We found that the aforementioned genes had dominance variance component estimates within the top 5% of all 17,994 gene expression probes that were analysed in ref. 22.

Functional enrichment analysis. Tissue specific transcriptionally active regions: we employed a recently published method (http://www.broadinstitute.org/mpg/epigwas/)[27] that tests for cell-type-specific enrichment of active chromatin, measured through H3K4me3 chromatin marks[27] in regions surrounding the 731 SNPs that comprise the 501 discovery interactions. The exact method used to perform this analysis has been described previously[37]. Briefly, we tested the hypothesis that the 731 SNPs were more likely to be in transcriptionally active regions (as measured by chromatin marks) than a random set of SNPs selected from the same SNP chip. This hypothesis was tested for 34 cell types across four broad tissue types (haematopoietic, gastrointestinal, musculoskeletal and endocrine, and brain).

Chromosome interactions: it has been shown[13] that different regions on different chromosomes or within chromosomes spatially co-localize within the cell. We shall refer to the co-localization of two chromosome regions as a chromosome interaction. A map of pairwise chromosome interactions for K562 blood cell lines was recently produced[29], and we speculated that part of the underlying biological mechanism behind some of the 501 epistatic interactions may arise from chromosome interactions. We found that 44 of the putative epistatic interactions were amongst SNPs that were within 5 Mb of known chromosome interactions. This means that SNP A was no more than 2.5 Mb from the focal point of the chromosome interaction on chromosome A, and SNP B was no more than 2.5 Mb from the focal point on chromosome B.

We performed simulations to test how extreme the observation of 44 epistatic interactions overlapping with chromosome interactions is compared to chance. Chromosome interactions fall within functional genomic regions[13,29], and the SNPs in our epistatic interactions are enriched for functional genomic regions. Therefore, we designed the simulations to ensure that the null distribution was of chromosome interactions between SNPs enriched for functional genomic regions but with no known epistatic interactions. To do this we used the 731 SNPs that form the 501 putative epistatic interactions and randomly shuffled them to create new sets of 501 pairs, disallowing any SNP combinations that were in the original set. Therefore, each new random set was enriched for functional regions but had no genetic interactions. We scanned the map of chromosome interactions for overlaps with the new sets and then repeated the random shuffling process. We performed 1,000 such permutations to generate a null distribution of chromosome interaction overlaps. We repeated this process, searching for overlaps within 1 Mb, 250 kb and 10 kb.

SNP co-localization with genomic features: we tested for enrichment of genomic features for the 687 IndexSNPs that comprise the 434 epistatic interactions with data present in discovery and replication data sets. For each of the 687 IndexSNPs we calculated linkage disequilibrium with all regional SNPs within a radius of 0.5 Mb

and kept all regional SNPs with linkage disequilibrium $r^2 > 0.8$. We then cross-referenced the remaining regional SNPs with the annotated chromatin structure reference[28]) querying whether the regional SNPs fell in predicted promoter region including TSS (TSS), predicted promoter flanking region (PF), predicted enhancer (E), predicted weak enhancer or open chromatin *cis* regulatory element (WE), CTCF enriched element (CTCF), predicted transcribed region (T) or predicted repressed or low activity region (R) positions. Therefore a particular IndexSNP might cover multiple genomic features through linkage disequilibrium.

We then performed the whole querying process for each of the 528,509 SNPs present in the SNP chip used in the scan and used the results from this second analysis to establish a null distribution for the expected proportion of SNPs for each genomic feature. We calculated *P* values for enrichment of each of the seven genomic features independently, and for *cis*- and *trans*-SNPs separately, using a binomial test. For each genomic feature we used the expected proportion of SNPs as the expected probability of 'success' (p). Here, a success is defined as an IndexSNP residing in a region that includes the genomic feature. The observed number of successes for each IndexSNP (k) out of the total count of IndexSNPs (n) was then modelled as $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Transcription factor enrichment: to test for enrichment of transcription factor binding sites (TFBS) we followed a procedure similar to that described in the previous section. For each of the 687 IndexSNPs we extracted regional SNPs as previously described. We then used the PWMEnrich package in Bioconductor (http://www.bioconductor.org/packages/2.12/bioc/html/PWMEnrich.html) to identify which TFBSs each of the regional SNPs for one IndexSNP falls in (within a radius of 250 bp). Thus, the number of occurrences of a particular TFBS was counted for each IndexSNP. We used the 'threshold-free affinity' method for identifying TFBSs[38].

We constructed a null distribution of expected TFBS occurrences based on the same null hypothesis as described in the previous section. The probability of an IndexSNP covering a particular TFBS is identical to any of the 528,509 SNPs in the discovery SNP chip. To do this, we performed the same procedure for each SNP in the discovery SNP chip as was performed for each IndexSNP to obtain an expected probability of covering a particular TFBS. We then tested the IndexSNPs for enrichment of each TFBS independently, and for *cis*- and *trans*-SNPs separately. *P* values were obtained using *z*-scores, calculated by using a normal approximation to the sum of binomial random variables representing motif hits along the sequence[39].

Defining SNP associations identified previously: the discovery data set (BSGS) had previously been analysed for additive and dominant marginal effects for all gene expression levels[22,23]. To define SNPs that had been previously detected to have effects for a particular gene expression level we used a significance threshold accounting for multiple testing across SNPs and expression probes, $T_m = 0.05/(528509 \times 7339) = 1.29 \times 10^{-11}$. From this, we found that only nine of the 501 discovery interactions had known main effects, 64 were between SNPs that had no known marginal effects, and 439 were between a SNP with a known marginal effect and a SNP with no known marginal effect.

31. Medland, S. E. *et al.* Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am. J Hum. Genet.* **85,** 750–755 (2009).
32. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23,** 1294–1296 (2007).
33. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genet.* **43,** 519–525 (2011).
34. Westra, H.-J. *et al.* MixupMapper: correcting sample mix-ups in genome-wide datasets increases power to detect small genetic effects. *Bioinformatics* **27,** 2104–2111 (2011).
35. Williams, D. A. Improved likelihood ratio tests for complete contingency tables. *Biometrika* **63,** 33–37 (1976).
36. Álvarez-Castro, J. M., Le Rouzic, A. & Carlborg, O. How to perform meaningful estimates of genetic effects. *PLoS Genet.* **4,** e1000062 (2008).
37. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340,** 1467–1471 (2013).
38. Stormo, G. D. DNA binding sites: representation and discovery. *Bioinformatics* **16,** 16–23 (2000).
39. Ho Sui, S. J. *et al.* oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res.* **33,** 3154–3164 (2005).

A synchronized global sweep of the internal genes of modern avian influenza virus

Michael Worobey[1], Guan-Zhu Han[1] & Andrew Rambaut[2,3]

Zoonotic infectious diseases such as influenza continue to pose a grave threat to human health[1]. However, the factors that mediate the emergence of RNA viruses such as influenza A virus (IAV) are still incompletely understood[2,3]. Phylogenetic inference is crucial to reconstructing the origins and tracing the flow of IAV within and between hosts[3–8]. Here we show that explicitly allowing IAV host lineages to have independent rates of molecular evolution is necessary for reliable phylogenetic inference of IAV and that methods that do not do so, including 'relaxed' molecular clock models[9], can be positively misleading. A phylogenomic analysis using a host-specific local clock model recovers extremely consistent evolutionary histories across all genomic segments and demonstrates that the equine H7N7 lineage is a sister clade to strains from birds—as well as those from humans, swine and the equine H3N8 lineage—sharing an ancestor with them in the mid to late 1800s. Moreover, major western and eastern hemisphere avian influenza lineages inferred for each gene coalesce in the late 1800s. On the basis of these phylogenies and the synchrony of these key nodes, we infer that the internal genes of avian influenza virus (AIV) underwent a global selective sweep beginning in the late 1800s, a process that continued throughout the twentieth century and up to the present. The resulting western hemispheric AIV lineage subsequently contributed most of the genomic segments to the 1918 pandemic virus and, independently, the 1963 equine H3N8 panzootic lineage. This approach provides a clear resolution of evolutionary patterns and processes in IAV, including the flow of viral genes and genomes within and between host lineages.

We speculate that distinct evolutionary rates in IAV lineages of different host species have compromised the accuracy of phylogenetic analyses with regard to inferring the timing and direction of the key movements of the virus between hosts over the last several centuries. To resolve this we have employed a model that allows IAV in individual host types to have different rates of molecular evolution, which we refer to as the host-specific local clock (HSLC) model. To evaluate its performance we generated synthetic nucleotide sequences along the model tree in Fig. 1a, using substitution model parameters and host-specific rates representative of real IAV data sets. Simulated 'equine', 'human' and 'avian' host clades were assigned relative rates of 1, 2 and 3, respectively, to cover the empirically observed range of rates (Fig. 2).

Figure 1 shows detailed results for the first replicate, but the patterns were almost identical across all replicates: a strict clock roots the tree incorrectly, resulting in the wrong maximum clade credibility (MCC) topology (posterior support = 1.0) and spurious timing estimates. A widely used relaxed clock model[9] also gets the topology wrong and very seriously underestimates the time to the most recent common ancestors (TMRCAs) at the deepest nodes. These models recovered the model tree topology in 0% of runs; the HSLC model did so in 100% of the simulations (Extended Data Fig. 1). Moreover, the relaxed clock's 95% credible interval for the TMRCA never included the real root node date; the HSLC model did in 91% of the simulations.

Hence, mis-specified molecular clock models are likely to generate profoundly misleading results with IAV data sets encompassing multiple host species. They are prone to severe systematic errors in topology and timing estimation, and erroneous conclusions can show strong statistical support. This is a serious problem given that such results are widely used to infer when, where and how pandemic and panzootic viruses have emerged.

Analyses of alignments of full-length segments encoding the IAV polymerase proteins (PB2, PB1 and PA), the haemagglutinin (HA)
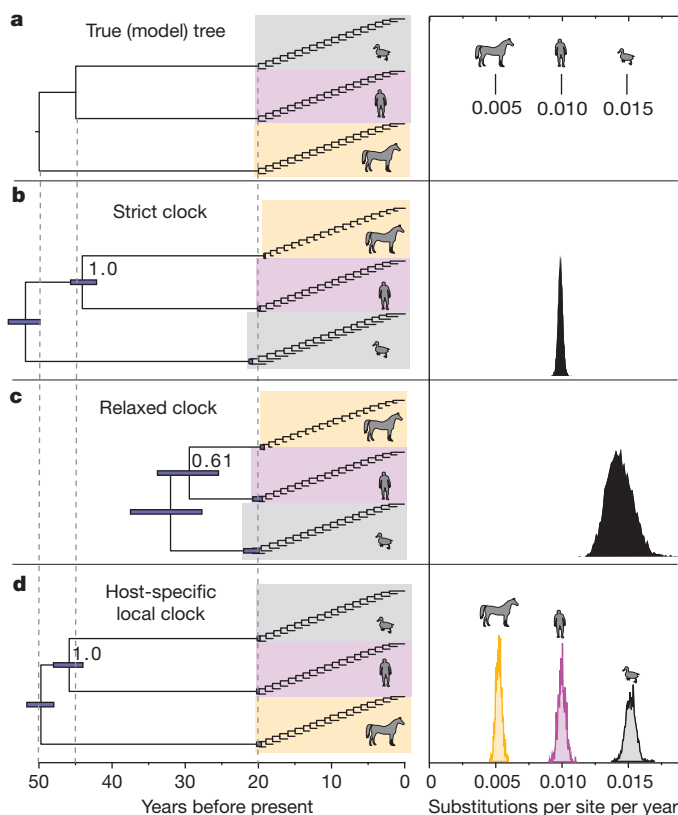

Figure 1 | Performance of different clock models on simulated data.
a, Model tree used to simulate nucleotide data, with branch lengths depicted in units of time. The host-specific rates of the 'equine', 'human' and 'avian' lineages are shown to the right of the tree. **b**, MCC tree (median node heights) for simulation replicate no. 1 under a strict clock model. The 95% credible interval for each node time is shown with a bar, and the posterior probability of the ingroup node is indicated. **c**, MCC tree (median node heights) under a relaxed clock model. **d**, MCC tree (median node heights) under the HSLC model. The posterior density of the clock rate inferred under each model is shown at the right. Summaries of the results for all 100 replicates are shown in Extended Data Fig. 1.

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA. [2]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. [3]Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA.
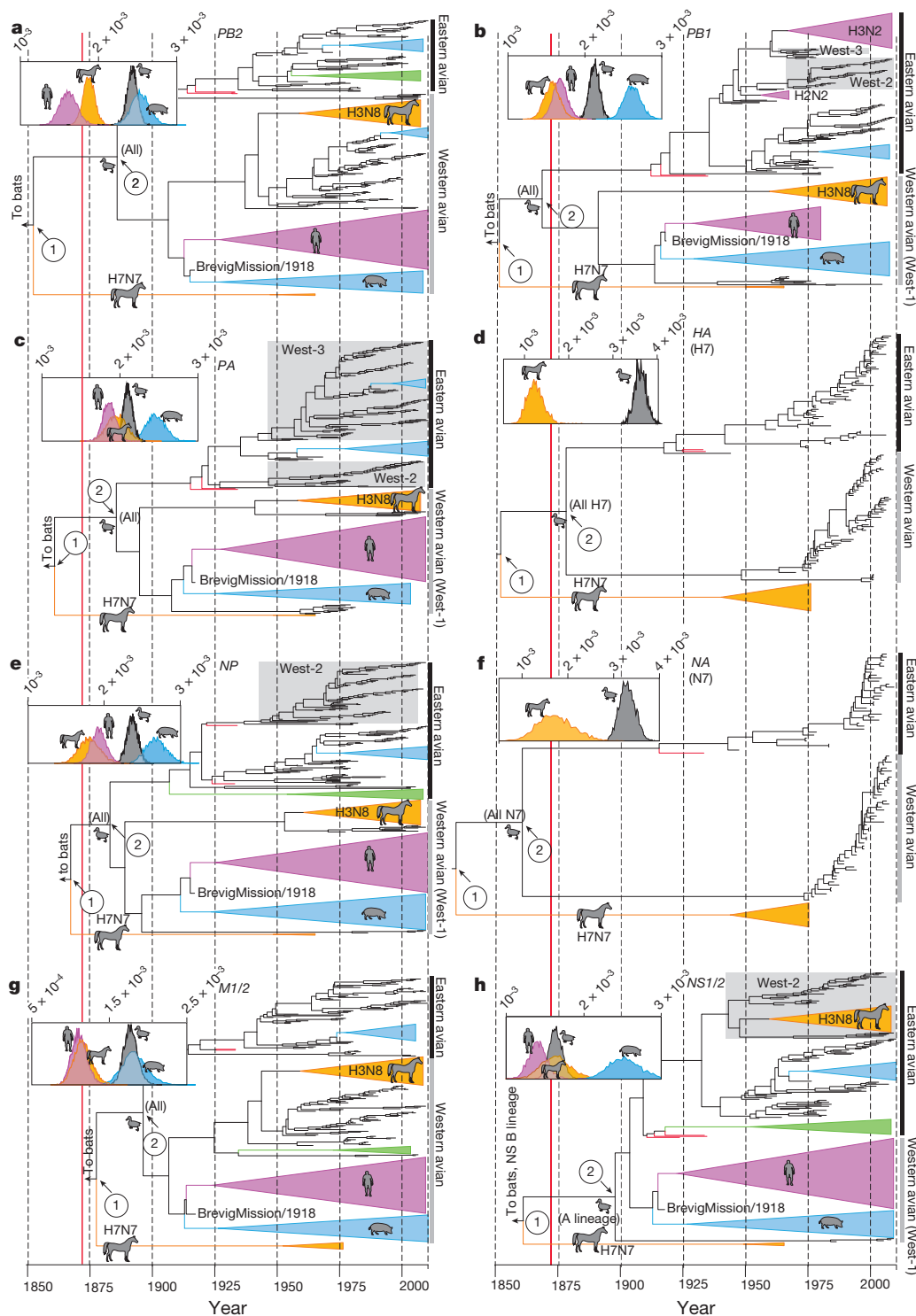
Figure 2 | Host-specific local clock model results. MCC trees (median node heights) and host-specific rate distributions (in substitutions per site per year, inset at top left of each panel) inferred under the HSLC model, for *PB2* (**a**), *PB1* (**b**), *PA* (**c**), *HA* (H7) (**d**), *NP* (**e**), *NA* (N7) (**f**), *M1/2* (**g**) and *NS1/2* (**h**). Trees are drawn to the same scale, with branch lengths in years. The epizootic of 1872–73 is indicated with a solid red line. The major eastern and western hemisphere AIV lineages are highlighted with black and grey vertical bars, respectively. The green triangles represent gull/shorebird clades (order Charadriiformes) allowed their own rate separate from other AIVs. In some cases (for example *NS1/2*) there is clear evidence of Charadriiformes AIV descending from domestic avian HPAI viruses of the 1920s and 1930s (which are highlighted in red). Fully resolved trees including posterior probabilities for each node and 95% credible intervals on node dates are depicted in Supplementary Fig. 1. Data, input and full MCC tree files are available from http://dx.doi.org/10.5061/dryad.m04j9.

surface glycoprotein subtypes H1, H3 and H7, the nucleocapsid protein (NP), the neuraminidase (NA) surface glycoprotein subtypes N1, N8 and N7, the matrix proteins (M1/2) and the nonstructural proteins (NS1/2) show how the HSLC model can outperform those assuming a single distribution of rates across hosts. In contrast with the conflicting topologies and timings inferred using a relaxed clock (Extended Data Fig. 2), the HSLC model revealed an underlying conformity and simplicity to IAV phylogenies (Fig. 2). A single pattern emerged, segment by segment: the equine H7N7 lineage[4,10] is positioned basal to all avian, human, swine and equine H3N8 strains (except for the *NS1/2* 'B' lineage,

discussed below). The timelines of the trees are also similar, with the most recent common ancestor (MRCA) of the equine H7N7 and avian lineages (node 1) dated between the 1830s and 1870s, and the MRCA of all avian strains (node 2) dated between the 1860s and 1890s.

Distinct western and eastern hemispheric avian clades are apparent, diverging at node 2, with more recent east-to-west introductions of the 'West-2' and 'West-3' clades in *PB1*, *PA*, *NP* and *NS1/2* (Fig. 2). Both node 1 and node 2 are supported by posterior probabilities of 1.0 in nearly every case (Supplementary Fig. 1). The results are robust to different substitution models and sampling regimes (Extended Data

Table 1), and analyses of third-codon position sites (Extended Data Table 1) and tests for episodic diversifying selection (Extended Data Fig. 3) suggest that they are unlikely to be biased by adaptive evolution.

For the branch between nodes 1 and 2 (Fig. 2) the period of overlap shared across all eight segments, accounting for uncertainty in node dates, is 1866–1878. This period is of interest because it encompasses one of the most extraordinary recorded outbreaks of influenza in animals, the severe western hemisphere panzootic of equine influenza in 1872–73 (refs 11, 12), which was also occasioned by widespread reports of influenza in domestic birds in the wake of local equine outbreaks[13]. The equine H7N7 lineage has an extraordinarily high uracil (U) content in all genes (Extended Data Figs 4 and 5). The U content of the IAV genome tends to increase steadily in mammalian hosts[14], so this is consistent with continuous circulation in horses from 1872 or perhaps even earlier. We speculate that there may have been some epidemiological event possibly associated with the 1872 epizootic that precipitated a global sweep of AIV internal genes, which we infer to have commenced around this time. The 1870s, moreover, saw the first scientific description of highly pathogenic avian influenza (HPAI) (in chickens in northern Italy[15]), which coincided with a transition to high-production, high-flock-size poultry farms[16].

Unlike previous approaches[17] (Extended Data Fig. 2), the HSLC model provides consistent, statistically strong evidence that the 1918 pandemic virus's *PB2*, *PB1*, *PA*, *NP*, *M1/2* and perhaps *NS1/2* arose from the western hemisphere AIV lineage (Fig. 2 and Supplementary Fig. 1). U content values are consistent with a recent avian origin of these internal genes, with none lying above the avian range in 1918 (Extended Data Fig. 4). The sampling of *PB1* and *PA* AIV sequences includes particularly close relatives of the 1918 virus, with which they shared a common ancestor just a few years before 1918 (Fig. 2). These suggest a North American origin of its internal genes, with domestic and wild birds being equally likely sources (Supplementary Fig. 1). These results mean that the hypothesis that several genes of the 1918 pandemic virus originated through a reassortment between human and swine lineages, circulating for decades in mammalian hosts[17], probably arises erroneously from a failure to accommodate differences in rates between hosts.

The internal proteins of the 1963 panzootic H3N8 equine virus also seem to have a western hemispheric avian origin (independent of the 1918 virus) (Fig. 2 and Extended Data Fig. 4). The best-resolved genes, *PA*, *NP* and *NS1/2*, show that this lineage shared an MRCA shortly before its emergence with avian strains from Argentina, Bolivia, Chile and Brazil (for example, *NP* avian/equine TMRCA: 1954 [1951–1957]). These findings are in agreement with the evidence that the virus, although first described in the United States, entered Miami, Florida, in 1963 through thoroughbred horses imported by air from Argentina[18].

When compared with the other genomic segments, the external antigenic genes *HA* and *NA* show considerably higher diversity in wild birds (Fig. 3) and no correlation in phylogenetic structure; see, for example, the H1, N1, H3 and N8 phylogenies in Extended Data Fig. 6. This has been attributed to a much higher rate of evolution in these genes, driven by antigenic adaptation, combined with free reassortment of the internal gene segments[7]. Our results suggest an intriguing alternative evolutionary pattern in which an avian virus, possibly an H7N7 from domestic birds, initiated a global selective sweep of its internal gene segments across the standing diversity of AIV, with the replacement of almost all previous AIV internal gene diversity. The exception is the *NS1/2* gene segment that was incompletely swept, leaving a rump of ancestral diversity (Supplementary Fig. 1h) designated the 'B' lineage[19]. We suspect that the sweep took several decades; an analysis of more recent succession dynamics in AIV (Fig. 2, Extended Data Table 2 and Supplementary Fig. 1) suggests that large geographic replacements are a continuing process and that typically several decades elapse between the emergence of a new variant and the completion of a global or hemispheric sweep.
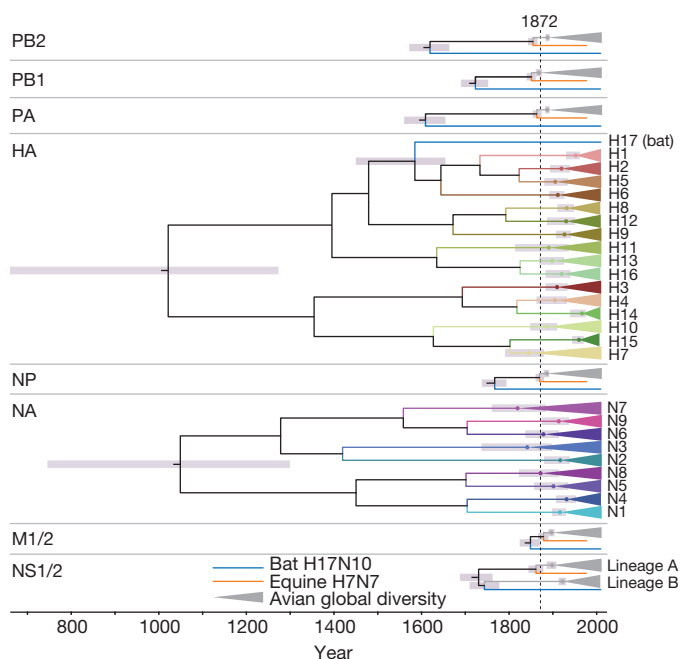

Figure 3 | HA, NA and internal gene diversity. Summarized time-calibrated phylogenetic trees of known IAV viruses for each genomic segment are shown. Each triangle represents global AIV diversity for internal genes (in grey) and each subtype of *HA* and *NA* (in colour). Grey bars represent 95% credible intervals for the dates of divergence of nodes of interest. The MRCA of each *HA* and *NA* subtype and the avian global diversity of every internal gene corresponds to or postdates 1872 (dashed line). Dates of divergence are shown in Extended Data Fig. 7c.

A sweep has been posited[8] as a possible explanation for the markedly restricted diversity of the internal genes in avian hosts (evident in Fig. 3), but that hypothesis was dismissed in favour of a model in which the internal genes are repeatedly hitchhiking on selective sweeps of the antigenic genes *HA* and *NA*. However, such a process could not have consistently removed all the diversity in the hitchhiked genes without the fixation of the selected gene as well. Indeed, genetic hitchhiking, by definition, involves a reduction of diversity in the locus that is the focus of selection[20], whereas here we observe *HA* and *NA* maintaining their diversity.

We propose that the global sweep of internal genes left some of the pre-existing diversity of *HA* and *NA* in place because immunity in previously exposed bird populations selects for antigenically novel viruses sufficiently to maintain multiple subtypes. Accordingly, phylogenies of *HA* and *NA* reveal the much older diversity of these gene segments: although evolving at different rates, *HA* and *NA* have indistinguishable TMRCAs of about 1,000 years ago (Fig. 3 and Extended Data Fig. 7). The TMRCA within each subtype is statistically compatible with a date of about 1870 or later, a result expected if the internal genes swept across standing diversity in *HA* and *NA* but difficult to explain otherwise. The current global diversity of AIV may thus descend from reassortment between these internal genes and pre-existing *HA*, *NA* and *NS1/2* B-lineage variants from as few as 16 avian viruses in the decades after initiation of the sweep (Fig. 3).

One lineage that did not receive this internal genome constellation is that found in fruit bats[21,22], presumably because avian-adapted genes did not confer sufficient fitness advantages or because the bat viruses are ecologically isolated. Under the model that the internal gene segment diversity was swept from avian (but not bat) influenza, the current position of the bat viruses as a basal lineage to the avian viral diversity in all genes except *HA*[21,22] and *NS1/2* (Supplementary Fig. 1) is readily explained without a reassortment event with an avian *HA*.

Descendants of 1920s–1930s H7 AIV from Eurasian domestic birds[16] account for virtually all eastern hemisphere AIV internal gene diversity

(Fig. 2 and Extended Data Fig. 8). Moreover, they were the source of *PB1*, *PA*, *NP* and *NS1/2* lineages that migrated between the 1930s and 1960s from the eastern hemisphere to the western hemisphere ('West-2' and 'West-3'; see Extended Data Table 2). By 2009–2013 these lineages had swept across the hemisphere, nearly displacing the older 'West-1' lineages (Fig. 2 and Supplementary Fig. 1). Global sweeps are thus still unfolding, and domestic avian influenza genes can displace previous variants—evidently even in wild hosts, and even on a hemispheric or global scale—over the course of decades. We suspect that domestic ducks may be a crucial nexus in the bidirectional (domestic/wild) gene flow of AIV worldwide (see Supplementary Information).

Influenza A viruses in humans frequently show global selective sweeps, with the entire H3N2 and H1N1 genomic diversity being replaced every few years[5,6]. During both the pandemics of the mid-twentieth century, new variants swept to fixation within months, completely replacing existing strains. In humans such dynamics are driven primarily by the fitness advantages that the virus obtains from antigenic novelty. However, the global avian selective sweep we have posited here was evidently driven by other selective advantages conferred by mutations in one or more of the internal genes. Characterizing the nature of these host-specific adaptations may be crucial in understanding the risk of future emergence of avian influenza in humans.

METHODS SUMMARY

We implemented the HSLC model in BEAST[23], allowing an arbitrary number of distinct clades to have rates inferred separately and simultaneously within a single Bayesian MCMC analysis. We constructed sequence alignments using all available full-segment IAV sequences encoding PB2, PB1, PA, HA (H1, H3, H7), NP, NA (N1, N7, N8), M1/2 and NS1/2 from birds, horses, pigs, humans and fruit bats. A subset of sequences of a size amenable to molecular clock analyses (about 300 sequences per segment) was sampled, preserving all major lineages and all the most basal sequences in major clades. Each major host group was allowed its own rate in the HSLC model. We inferred the rate of increase of U content through time for equine H3N8 by fitting a two-dimensional Scaled Log Transform: U content = $a \ln(b \times \text{year} + c)$. Age estimates for equine H7N7 were calculated by assuming these rates and were then compared with the timing of the 1872 epizootic to test whether the high U content in equine H7N7 is consistent with continuous equine transmission from this period.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Morens, D. M., Folkers, G. K. & Fauci, A. S. The challenge of emerging and re-emerging infectious diseases. *Nature* **430,** 242–249 (2004).
2. Parrish, C. R. *et al.* Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. Rev.* **72,** 457–470 (2008).
3. Holmes, E. C. *The Evolution and Emergence of RNA Viruses* (Oxford Univ. Press, 2009).
4. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56,** 152–179 (1992).
5. Fitch, W. M., Bush, R. M., Bender, C. A. & Cox, N. J. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl Acad. Sci. USA* **94,** 7712–7718 (1997).
6. Rambaut, A. *et al.* The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453,** 615–619 (2008).
7. Dugan, V. G. *et al.* The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog.* **4,** e1000076 (2008).
8. Chen, R. & Holmes, E. C. Hitchhiking and the population genetic structure of avian influenza virus. *J. Mol. Evol.* **70,** 98–105 (2010).
9. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4,** e88 (2006).
10. Sovinova, O., Tumova, B., Poutska, F. & Nemec, J. Isolation of a virus causing respiratory disease in horses. *Acta Virol.* **2,** 52–61 (1958).
11. Morens, D. M. & Taubenberger, J. K. Historical thoughts on influenza viral ecosystems, or behold a pale horse, dead dogs, failing fowl, and sick swine. *Influenza Other Respir. Viruses* **4,** 327–337 (2010).
12. Judson, A. B. History and course of the epizoötic among horses upon the North American continent in 1872–73. *Publ. Health Pap. Rep.* **1,** 88–109 (1873).
13. Morens, D. M. & Taubenberger, J. K. An avian outbreak associated with panzootic equine influenza in 1872: an early example of highly pathogenic avian influenza? *Influenza Other Respir. Viruses* **4,** 373–377 (2010).
14. Rabadan, R., Levine, A. J. & Robins, H. Comparison of avian and human influenza A viruses reveals a mutational bias on the viral genomes. *J. Virol.* **80,** 11887–11891 (2006).
15. Perroncito, E. Epizoozia tifoide nei gallinacei. *Annli Reale Accad. Agric. Torino* **21,** 87–126 (1878).
16. Kaleta, E. F. & Rülke, C. P. A. in *Avian Influenza* (ed. Swayne, D. E.) 145–189 (Blackwell, 2008).
17. Smith, G. J. D. *et al.* Dating the emergence of pandemic influenza viruses. *Proc. Natl Acad. Sci. USA* **106,** 11709–11712 (2009).
18. Scholtens, R. G. & Steele, J. H. U. S. epizootic of equine influenza, 1963: Epizootiology. *Public Health Rep.* **79,** 393–398 (1964).
19. Treanor, J. J., Snyder, M. H., London, W. T. & Murphy, B. R. The B allele of the NS gene of avian influenza viruses, but not the A allele, attenuates a human influenza A virus for squirrel monkeys. *Virology* **171,** 1–9 (1989).
20. Barton, N. H. Genetic hitchhiking. *Phil. Trans. R. Soc. Lond. B* **355,** 1553–1562 (2000).
21. Tong, S. *et al.* A distinct lineage of influenza A virus from bats. *Proc. Natl Acad. Sci. USA* **109,** 4269–4274 (2012).
22. Tong, S. *et al.* New World bats harbor diverse influenza A viruses. *PLoS Pathog.* **9,** e1003657 (2013).
23. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–1973 (2012).

Author Contributions M.W., G.-Z.H. and A.R. designed the study. M.W. and A.R. conceived the analytical approach, and A.R. developed the software. G.-Z.H., M.W. and A.R. prepared the data sets. M.W., G.-Z.H. and A.R. performed the phylogenetic analyses. M.W. conducted the U content analyses. M.W. and A.R. wrote the paper. All authors discussed all the results and commented on the manuscript.

Author Information Sequences for A/equine/Detroit/3/1964(H7N7), A/chicken/Japan/1925(H7N7) and A/duck/Manitoba/1953(H10N7) have been deposited in the GenBank database under accession numbers KF435047–KF435062 and KF619244–KF619250. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W. (worobey@email.arizona.edu) or A.R. (a.rambaut@ed.ac.uk).

LETTER

doi:10.1038/nature13198

Nicotinamide *N*–methyltransferase knockdown protects against diet–induced obesity

Daniel Kraus[1]*†, Qin Yang[1]*†, Dong Kong[1], Alexander S. Banks[2], Lin Zhang[1], Joseph T. Rodgers[2], Eija Pirinen[3]†, Thomas C. Pulinilkunnil[1]†, Fengying Gong[1]†, Ya–chin Wang[1], Yana Cen[4], Anthony A. Sauve[4], John M. Asara[5], Odile D. Peroni[1], Brett P. Monia[6], Sanjay Bhanot[6], Leena Alhonen[3]†, Pere Puigserver[2] & Barbara B. Kahn[1]

In obesity and type 2 diabetes, *Glut4* glucose transporter expression is decreased selectively in adipocytes[1]. Adipose-specific knockout or overexpression of *Glut4* alters systemic insulin sensitivity[2]. Here we show, using DNA array analyses, that nicotinamide *N*-methyltransferase (*Nnmt*) is the most strongly reciprocally regulated gene when comparing gene expression in white adipose tissue (WAT) from adipose-specific *Glut4*-knockout or adipose-specific *Glut4*-overexpressing mice with their respective controls. NNMT methylates nicotinamide (vitamin B3) using *S*-adenosylmethionine (SAM) as a methyl donor[3,4]. Nicotinamide is a precursor of NAD$^+$, an important cofactor linking cellular redox states with energy metabolism[5]. SAM provides propylamine for polyamine biosynthesis and donates a methyl group for histone methylation[6]. Polyamine flux including synthesis, catabolism and excretion, is controlled by the rate-limiting enzymes ornithine decarboxylase (ODC) and spermidine–spermine N^1-acetyltransferase (SSAT; encoded by *Sat1*) and by polyamine oxidase (PAO), and has a major role in energy metabolism[7,8]. We report that NNMT expression is increased in WAT and liver of obese and diabetic mice. *Nnmt* knockdown in WAT and liver protects against diet-induced obesity by augmenting cellular energy expenditure. NNMT inhibition increases adipose SAM and NAD$^+$ levels and upregulates ODC and SSAT activity as well as expression, owing to the effects of NNMT on histone H3 lysine 4 methylation in adipose tissue. Direct evidence for increased polyamine flux resulting from NNMT inhibition includes elevated urinary excretion and adipocyte secretion of diacetylspermine, a product of polyamine metabolism. NNMT inhibition in adipocytes increases oxygen consumption in an ODC-, SSAT- and PAO-dependent manner. Thus, NNMT is a novel regulator of histone methylation, polyamine flux and NAD$^+$-dependent SIRT1 signalling, and is a unique and attractive target for treating obesity and type 2 diabetes.

Nnmt is expressed at high levels in adipose tissue and liver and at lower levels in other organs[3,4]. *Nnmt* is increased in multiple cancers[9,10], neurodegenerative diseases[11], and also in obesity and diabetes[12–14]. For example, *Nnmt* expression is increased in adipocytes of obese, compared to non-obese Pima Indians[12]. Metabolomic analyses reveal elevated levels of urinary N^1-methylnicotinamide, the product of NNMT, in humans with type 2 diabetes, and in *db/db* mice and obese Zucker rats[13], indicating increased NNMT activity in obesity and type 2 diabetes. Quantitative trait loci mapping in mice suggests a causative role of *Nnmt* in type 2 diabetes[14]. Therefore, we investigated the regulation and roles of NNMT in obesity and type 2 diabetes.

Adipose *Nnmt* messenger RNA levels were increased twofold in insulin-resistant adipose-specific *Glut4*-knockout mice (Fig. 1a), and reduced by 62% in insulin-sensitive adipose-specific *Glut4*-overexpressing mice (Fig. 1b). NNMT protein was increased 1.5- to 2-fold in WAT of *ob/ob*, *db/db* mice and high-fat diet (HFD)-fed mice (Fig. 1c–e), compared with lean, insulin-sensitive controls. Hepatic NNMT protein levels were increased in *ob/ob* and *db/db* mice (Fig. 1f, g) and tended to be higher in HFD-fed mice (Fig. 1h). Thus, NNMT is upregulated in adipose tissue and liver of mouse models of obesity and insulin resistance.

Adipose and hepatic *Nnmt* expression varies highly among 25 different mouse strains[15] (Supplementary Fig. 1). Adipose *Nnmt* expression is high in obesity-prone strains and low in obesity-resistant strains[15–17] (Supplementary Fig. 1a). In contrast, liver *Nnmt* expression does not parallel the propensity for obesity (Supplementary Fig. 1b). Adipose *Nnmt* expression correlates highly with per cent fat mass in diet-induced obesity across 20 different mouse strains[18], and with expression of retinol-binding protein 4 (*Rbp4*) ($r = 0.90$, $P < 0.0001$ by Pearson correlation coefficient, two-tailed test), an adipokine that contributes to insulin resistance[2] (Supplementary Fig. 2a, b). Thus, we proposed that elevated NNMT levels in adipose tissue and/or liver may have a causative role in insulin resistance and obesity.

We knocked down *Nnmt* with antisense oligonucleotides (ASOs) in HFD-fed mice. ASOs regulate gene expression primarily in liver and fat[19,20]. Treating HFD-fed mice with *Nnmt* ASO reduced *Nnmt* mRNA and NNMT protein by 60 to 75% in WAT and 60% (protein) to 90% (mRNA) in liver (Fig. 2a–d), but not in brown adipose tissue and kidney (Fig. 2e, f). Serum transaminases and creatinine were normal, indicating no hepatic or renal toxicity with ASO treatment (Supplementary Fig. 3a–c).

Nnmt knockdown in adipose tissue and liver protected mice from diet-induced obesity (Fig. 2g), causing a 47% reduction in relative fat mass and a 15% increase in relative lean mass (Fig. 2h, i). Free body water and hepatic glycogen content were not different (not shown) between control- and *Nnmt*-ASO-treated mice. Subcutaneous and epididymal fat-pad weights were lower in *Nnmt*-knockdown mice than in controls (Fig. 2j, k), largely owing to reduced adipocyte size (44% reduced area, Fig. 2l, and 70% reduced volume, Supplementary Fig. 4). Insulin sensitivity was enhanced, as evidenced by approximately 50 to 60% lower serum insulin levels and glucose-insulin product (Fig. 2m, n). *Nnmt* knockdown also improved glucose tolerance (Fig. 2o, p), prevented HFD-induced hepatic steatosis (Fig. 2q, r), and decreased serum triglycerides and free fatty acids (Supplementary Fig. 5a, b). Thus, *Nnmt* knockdown in WAT and liver protects against diet-induced obesity and its deleterious metabolic consequences.

[1]Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, 330 Brookline Avenue, Boston, Massachusetts 02215, USA. [2]Department of Cancer Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Biotechnology and Molecular Medicine, A.I. Virtanen Institute for Molecular Sciences, Biocenter Kuopio, University of Eastern Finland, Kuopio Campus, PO Box 1627, FI-70211 Kuopio, Finland. [4]Department of Pharmacology, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10065, USA. [5]Division of Signal Transduction, Beth Israel Deaconess Medical Center and Harvard Medical School, 330 Brookline Ave, Boston, Massachusetts 02215, USA. [6]Isis Pharmaceuticals, 1896 Rutherford Road, Carlsbad, California 92008-7326, USA. †Present addresses: Division of Nephrology, Department of Internal Medicine I, Würzburg University Hospital, Oberdürrbacher Straße 6, 97080 Würzburg, Germany (D.K.); Department of Medicine, Physiology and Biophysics, Center for Diabetes Research and Treatment, and Center for Epigenetics and Metabolism, University of California, Irvine, California 92697, USA (Q.Y.); Research Programs Unit, Molecular Neurology, Biomedicum Helsinki, University of Helsinki, 00290, Helsinki, Finland (E.P.); Department of Biochemistry and Molecular Biology, Faculty of Medicine, Dalhousie Medicine New Brunswick, Dalhousie University, Saint John, New Brunswick E2L4L5, USA (T.C.P.); Department of Endocrinology, Key Laboratory of Endocrinology of Ministry of Health, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China (F.G.); School of Pharmacy, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Finland (L.A.) .
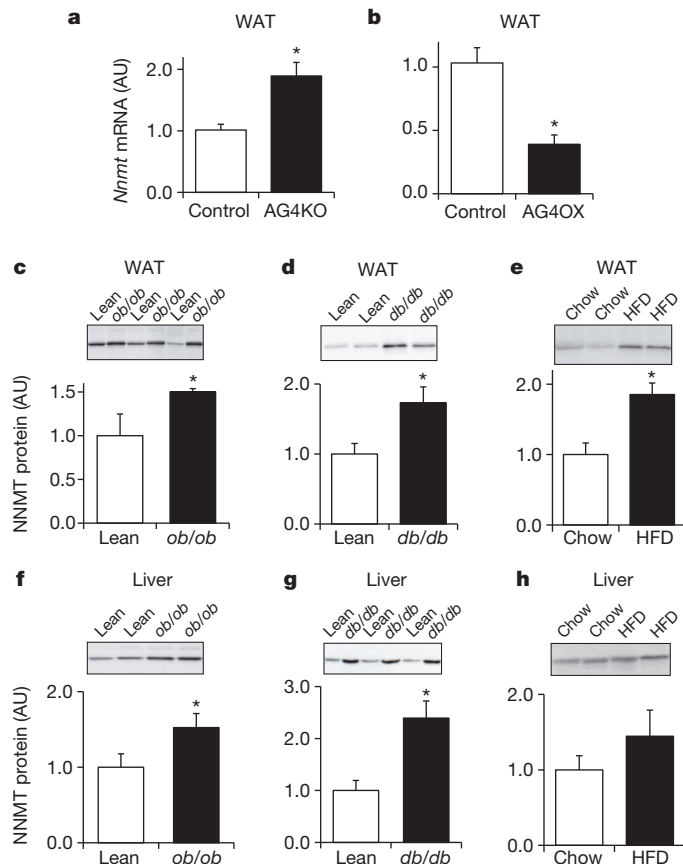*These authors contributed equally to this work.

Figure 1 | NNMT expression is increased in WAT and liver of obese and insulin-resistant mice. a, b, *Nnmt* mRNA expression normalized by cyclophilin in WAT of adipose-specific *Glut4* knockout (AG4KO) mice and aP2-Cre controls (*n* = 4 per group) (a) and adipose-specific *Glut4* overexpressing (AG4OX) and wild-type littermate controls (*n* = 6 per group) (b). c–e, NNMT protein levels in WAT of *ob/ob* mice (*n* = 8) and lean controls (*n* = 4) (c); *db/db* mice and lean controls (*n* = 7 per group) (d), and high-fat diet (HFD)-fed (*n* = 6) and chow-fed mice (*n* = 7) (e). f–h, NNMT protein levels in liver of *ob/ob* mice (*n* = 9) and lean controls (*n* = 6) (f); *db/db* mice and lean controls (*n* = 7 per group) (g); and HFD-fed and chow-fed mice (*n* = 6 per group) (h). Actin was used as a control for western blot analysis and the levels were not different between lean and obese mice. AU, arbitrary units. Error bars, ± s.e.m; *P < 0.05.

We sought to determine whether the leanness with *Nnmt* knockdown is due to reduced energy intake or increased energy expenditure. The NNMT substrate, nicotinamide, at pharmacological doses may suppress food intake and cause weight loss in rats[21]. However, *Nnmt*-ASO-treated mice consumed the same amount of calories as controls (Fig. 3a). Food intake was measured before body weights diverged, in order to eliminate confounding by body-weight differences. Feed efficiency (body weight change per kcal of food eaten) was reduced by approximately 50% in *Nnmt*-knockdown mice (Fig. 3b). Although control-ASO-treated mice gained about 7 mg of fat per kcal of food intake, *Nnmt*-ASO-treated mice gained no fat (Fig. 3c). Thus, body weight and fat gain were disproportionally low for the calorie intake in *Nnmt*-ASO-treated compared with control-ASO-treated mice. Faecal lipid excretion was not changed between control- and *Nnmt*-ASO-treated mice (Fig. 3d). Therefore, increased energy expenditure rather than decreased food intake or steatorrhoea explains the leanness in *Nnmt*-knockdown mice. This was further demonstrated by increased oxygen consumption (expressed per kg body weight or per mouse) during comprehensive laboratory animal monitoring system (CLAMS) analysis performed before body weights started to diverge (Fig. 3e and Supplementary Fig. 6a). Comparison of mice with similar body weight showed that *Nnmt*-ASO-treated mice had higher energy expenditure than control-ASO-treated mice (Supplementary
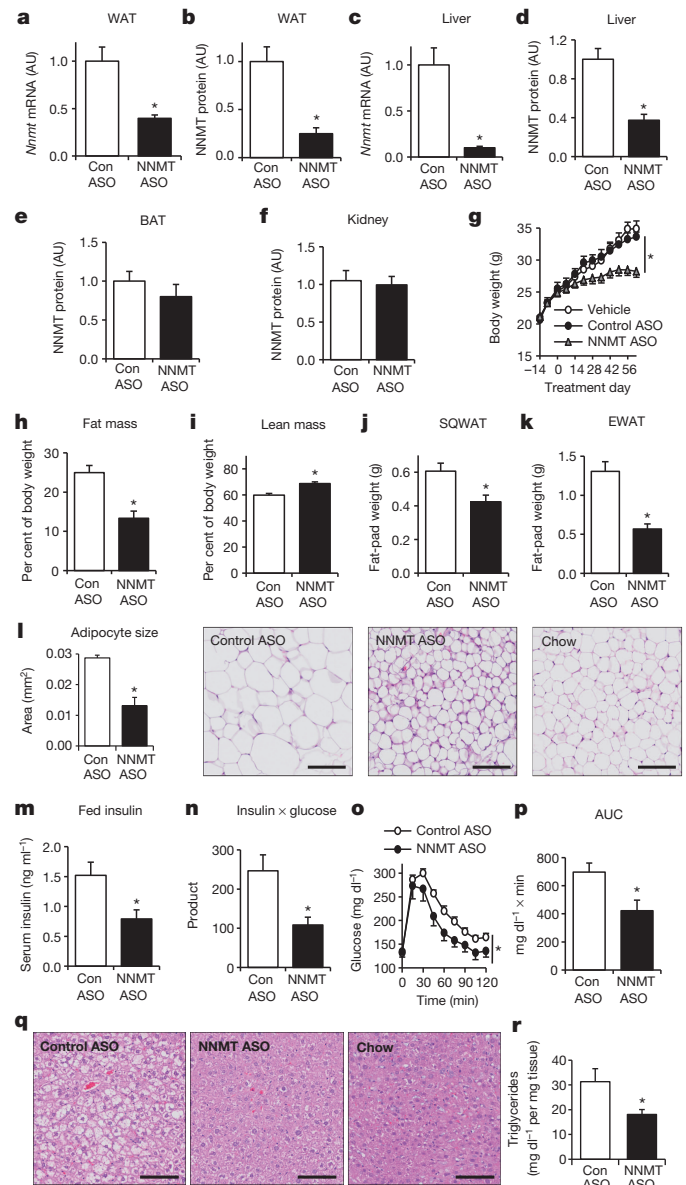

Figure 2 | *Nnmt* knockdown prevents diet-induced obesity and insulin resistance. a–f, Knockdown efficiency of *Nnmt*-ASO. mRNA expression was normalized by cyclophilin and protein levels were corrected with actin levels: *Nnmt* mRNA (a) and NNMT protein in WAT (b); *Nnmt* mRNA (c) and NNMT protein in liver (d); NNMT protein in brown adipose tissue (BAT) (e) and kidney (f). g, Body weights of C57BL/6 mice fed a high-fat diet and treated with *Nnmt* ASO, control ASO, or vehicle (saline) for 8 weeks. h, Fat mass as a percentage of body weight. i, Lean mass as a percentage of body weight. j, Subcutaneous WAT (SQWAT) fat-pad weights. k, Epididymal WAT (EWAT) fat-pad weights. l, Epididymal adipocyte cross-sectional area and haematoxylin and eosin (H&E)-stained sections of SQWAT. m, Serum insulin levels. n, Glucose × insulin product (ng ml⁻¹ × mg dl⁻¹) in the fed state. o, Intraperitoneal glucose tolerance test. p, Area under the curve (AUC) of the glucose tolerance. q, H&E stain of liver sections of HFD-fed *Nnmt*-ASO- and control-ASO-treated mice, and of chow-fed mice. r, Hepatic triglyceride levels in *Nnmt*- and control-ASO-treated mice. The scale bars in l and q represent 100 μm; *n* = 8 per group for a–p, *n* = 13 per group for r. AU, arbitrary units. Error bars, ± s.e.m.; *P < 0.05.

Fig. 6b). Energy expenditure per gram of fat mass, but not per gram of lean mass was increased in *Nnmt*-ASO-treated mice compared with control-ASO-treated mice (Supplementary Fig. 6c, d). There was no measurable difference in food intake expressed per mouse or per g body weight, locomotor activity or respiratory exchange ratio (RER) (Supplementary Fig. 7a–d). As a third approach to assess the energy
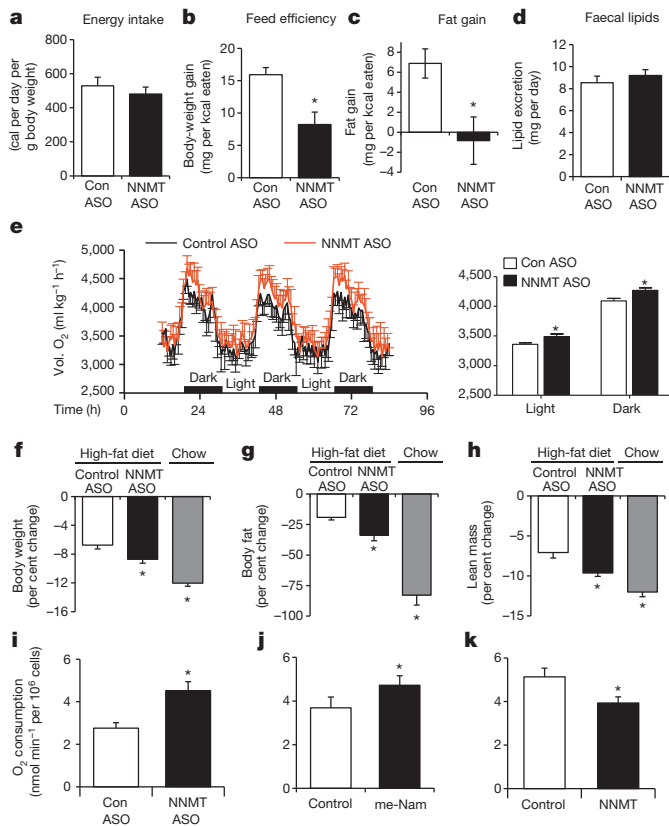
Figure 3 | NNMT regulates energy expenditure. a, Energy intake of HFD-fed control-ASO- and *Nnmt*-ASO-treated mice. **b**, Feed efficiency (body-weight gain per kilocalorie (kcal) consumed) of HFD-fed control-ASO- and *Nnmt*–ASO-treated mice. **c**, Fat mass gain per kcal consumed in HFD-fed mice treated with *Nnmt* ASO or control ASO. $n = 12$ per group for **a–c**. **d**, Faecal lipid excretion ($n = 10$ per group). **e**, Oxygen consumption (Vol. O_2) measured by CLAMS ($n = 7$ per group). **f–h**, Effects of a 16-h fast on body weight (**f**), fat (**g**), and lean body mass (**h**). **i–k**, Oxygen consumption in 3T3-L1 adipocytes transfected with control or *Nnmt* ASO ($n = 6$ per group) (**i**), treated with 10 mM N^1-methylnicotinamide (me-Nam) ($n = 6$ per group) (**j**), or transfected with *Nnmt* cDNA ($n = 10$ per group) (**k**). Error bars, ± s.e.m.; *$P < 0.05$.

expenditure, we measured body weight and body composition before and after an overnight fast. In the absence of energy intake, increased loss of body fat indicates increased energy expenditure. After fasting, *Nnmt*-knockdown mice on HFD lost more body weight, body fat and lean mass than control-ASO-treated mice, and the response resembled that of chow-fed lean mice (Fig. 3f–h). Core body temperature, BAT weight, and uncoupling protein-1 levels in BAT (Supplementary Fig. 8a–c) were unchanged. Thus, *Nnmt* knockdown in WAT and liver increases energy expenditure, and this is likely to be independent of BAT-induced energy dissipation or heat production.

We asked whether NNMT regulates energy expenditure in a cell-autonomous manner. *Nnmt* knockdown in adipocytes led to a 60% increase in oxygen consumption (Fig. 3i). NNMT inhibition by N^1-methylnicotinamide—the NNMT reaction product and a specific, potent inhibitor of NNMT at pharmacological doses[3]—also increased oxygen consumption (Fig. 3j). Conversely, *Nnmt* overexpression decreased oxygen consumption (Fig. 3k) in adipocytes. Similar results were obtained in cultured hepatoma cells (Supplementary Fig. 9). Therefore, NNMT regulates oxygen consumption in a cell-autonomous fashion in both adipocytes and hepatocytes.

We sought to determine the molecular mechanisms for increased energy expenditure with NNMT inhibition. NNMT methylates nicotinamide using SAM as a methyl donor and generates *S*-adenosylhomocysteine (SAH). Adipose SAM, and the SAM:SAH ratio were increased by 50%

in *Nnmt*-ASO-treated mice (Fig. 4a). In liver, SAM was not changed, but the SAM:SAH ratio was increased by 2.2-fold owing to a 48% reduction of SAH levels in *Nnmt*-ASO-treated mice (Supplementary Fig. 10). The unchanged hepatic SAM with *Nnmt* knockdown contrasts with knockdown of glycine *N*-methyltransferase (*Gnmt*), a dominant methyltransferase that accounts for >1% of total cytosolic protein in liver. *Gnmt* knockdown increases SAM by 40-fold and induces hepatic steatosis and fibrosis[22]. Hepatic *Nnmt* knockdown did not cause fibrosis as evidenced by normal expression of tissue inhibitor of metalloproteinase 1 (*Timp1*) and collagen type 1 (*Col1a1*), markers for hepatic fibrosis (Supplementary Fig. 11a, b).

SAM has two major functions: first, providing propylamine groups for polyamine biosynthesis; and second, donating methyl groups to substrates including histones[6]. Polyamines (putrescine, spermine and spermidine) are organic polycations that are essential for multiple cellular functions affecting cell growth, cancer and ageing[23]. Polyamine metabolism is tightly controlled (Supplementary Fig. 12)[23]. Synthesis is controlled by ODC producing putrescine, and by adenosylmethionine decarboxylase (AMD1) providing decarboxylated SAM for the synthesis of spermidine and spermine. Catabolism is controlled by spermidine–spermine N^1-acetyltransferase (SSAT), which acetylates spermidine and spermine using acetyl-CoA as a substrate. The acetylated products including N^1-acetylspermine, N^1,N^{12}-diacetylspermine, and N^1-acetylspermidine, are either oxidized by PAO or excreted intact in urine. Polyamine flux has a major role in energy homeostasis. Genetic knockout of SSAT results in increased diet-induced obesity and transgenic SSAT overexpression causes leanness owing to altered energy expenditure[7,8,24]. As *Nnmt* knockdown increases adipose SAM levels (Fig. 4a), we reasoned that this may cause a 'substrate shunt' of SAM from the NNMT reaction to polyamine flux leading to increased energy expenditure in *Nnmt*-ASO-treated mice (Fig. 3). In support of this, *Nnmt* expression correlates negatively with *Odc* and *Ssat* expression in adipose tissue, but not in liver in 25 different mouse strains[15] (Supplementary Fig. 13a–d).

Nnmt-ASO treatment *in vivo* augmented adipose ODC and SSAT activity (Fig. 4b, c) and mRNA expression (Fig. 4d). The expression of adenosylmethionine decarboxylase *Amd1*, which provides decarboxylated SAM for spermidine and spermine synthesis, was also increased (Fig. 4d). In liver, *Nnmt* ASO increased ODC, but not SSAT activity, or *Odc* or *Ssat* expression (Supplementary Fig. 14a–d). ODC and SSAT activation drives polyamine flux, which consumes the metabolic substrate, acetyl-CoA, for polyamine acetylation[7]. Consistent with reduced acetyl-CoA availability for lipogenesis, acetyl-CoA carboxylase 1 expression and fatty acid synthase expression and activity were decreased in adipose tissue of *Nnmt*-ASO-treated mice (Supplementary Fig. 15a–c). In addition, adipose *Nnmt* knockdown also decreased ATP, increased the AMP:ATP ratio and enhanced AMPK threonine 172 phosphorylation (Supplementary Fig. 16a–d). Increased urinary diacetylspermine provides direct evidence for enhanced adipose polyamine flux in *Nnmt*-ASO-treated mice (Fig. 4e). Furthermore, reduction in *Nnmt* increased *Odc* and *Ssat* expression in cultured adipocytes (Fig. 4f, g). N^1-methylnicotinamide treatment promoted diacetylspermine secretion from adipocytes in a dose-dependent manner (Fig. 4h), complementing the increased urinary diacetylspermine excretion in *Nnmt*-ASO-treated mice (Fig. 4e). These data convincingly show that NNMT directly regulates polyamine flux in adipocytes.

In addition to being a substrate for polyamine metabolism, SAM is a methyl donor for numerous methylation reactions including histone methylation, which is important for transcriptional regulation[6]. We measured eight types of histone methylation; the amounts of mono-, di- or tri-methylated lysine 4 of histone H3 (H3K4) were increased in adipose tissue of *Nnmt*-ASO-treated mice (Fig. 4i). Chromatin immunoprecipitation followed by quantitative PCR (ChIP–qPCR) revealed enrichment of methylated H3K4 on *Odc* and *Ssat* genes with NNMT inhibition in adipocytes (Fig. 4j). Thus, NNMT inhibition modifies histone methylation and increases *Odc* and *Ssat* expression, leading to activation of polyamine flux.
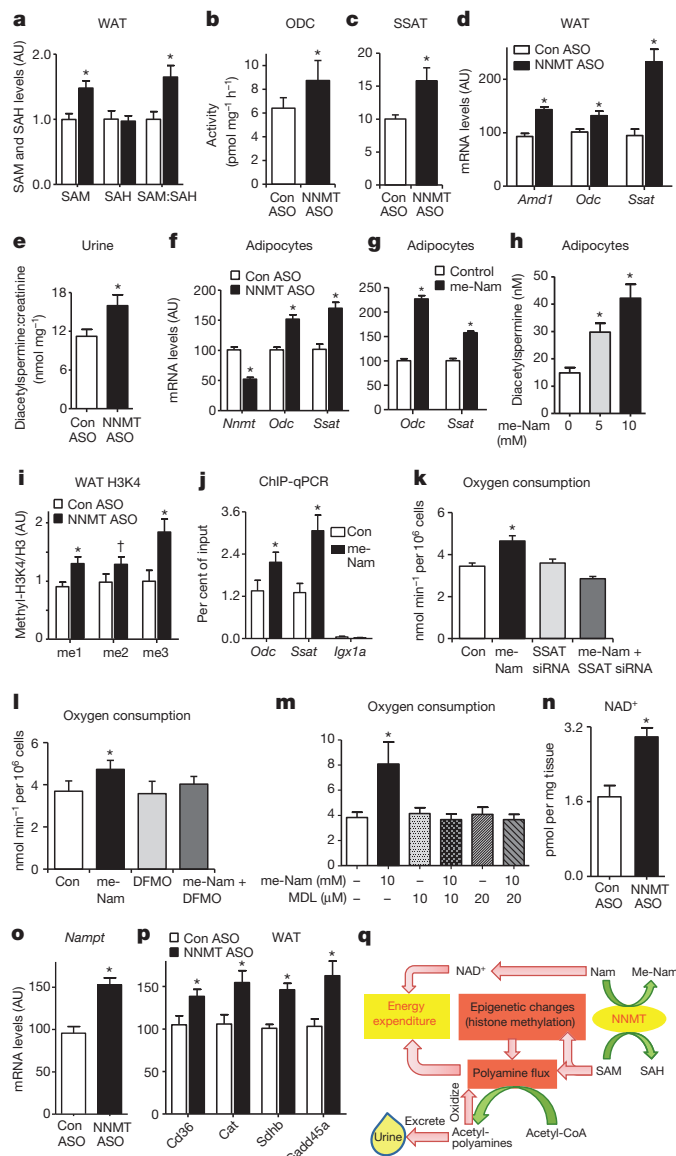
Figure 4 | NNMT regulates SAM and NAD⁺ pathways in adipose tissue.
a, Adipose S-adenosylmethionine (SAM), S-adenosylhomocysteine (SAH) and SAM:SAH ratio measured by LC–MS/MS ($n = 8$ for control-ASO-treated mice; $n = 12$ for $Nnmt$-ASO-treated mice). **b**, **c**, ODC (**b**) and SSAT (**c**) activity ($n = 10$ per group). **d**, $Amd1$, Odc and $Ssat$ mRNA expression in adipose tissue of $Nnmt$-ASO- and control-ASO-treated mice ($n = 12$ per group). **e**, Urinary diacetylspermine:creatinine ratio ($n = 22$ for control-ASO-treated mice; $n = 29$ for $Nnmt$-ASO-treated mice). **f**, **g**, Odc and $Ssat$ mRNA levels in 3T3-L1 adipocytes with $Nnmt$ knockdown ($n = 9$ per group) (**f**) and N^1-methylnicotinamide (me-Nam) (**g**) treatment ($n = 6$ per group). **h**, Diacetylspermine secretion from 3T3-L1 adipocytes treated with me-Nam ($n = 10$ per group). **i**, Expression of mono-, di- and tri-methylation of lysine histone 3 (H3K4) normalized to total H3 levels in adipose tissue ($n = 8$ per group). **j**, H3K4me2 occupancy on Odc, $Ssat$ genes and an open reading frame free region ($Igx1a$) as a negative control in adipocytes measured by ChIP-qPCR ($n = 12$ per group). **k–m**, Oxygen consumption in adipocytes transfected with control or $Ssat$ siRNA, and treated with or without 10 mM me-Nam ($n = 6$ per group) (**k**); treated with 10 mM me-Nam with or without 5 mM difluoromethylornithine (DFMO), a specific ODC inhibitor ($n = 6$ per group) (**l**); or treated with 10 mM me-Nam with or without MDL72527, a specific PAO inhibitor ($n = 6$ per group) (**m**). **n**, NAD⁺ levels ($n = 7$ Con-ASO; $n = 12$ $Nnmt$ ASO). **o**, mRNA levels of nicotinamide phosphoribosyltransferase ($Nampt$) ($n = 11$ per group). **p**, mRNA levels of SIRT1 target genes: $Cd36$, catalase (Cat), succinate dehydrogenase B ($SdhB$) and growth arrest and DNA-damage-inducible protein ($Gadd45a$) in adipose tissue of NNMT-ASO- and control-ASO treated mice ($n = 11$ per group). Error bars, $\pm$ s.e.m., *$P < 0.05$, †$P = 0.06$. **q**, Model of NNMT-regulated energy expenditure in adipocytes. NNMT methylates nicotinamide (Nam), a precursor of NAD⁺, using SAM as a methyl donor. SAM regulates polyamine flux by providing substrates and modulating histone methylation. Polyamine flux utilizes acetyl-CoA to generate acetyl-polyamines, which are oxidized or excreted in the urine. Taken together, this results in adipose metabolic substrate consumption and loss coupled with systemic alteration of energy expenditure.

$Nnmt$-ASO-treated mice are leaner, NAD⁺ levels per adipocyte may not be as elevated as per mg of adipose tissue. However, the increased gene expression upstream and downstream of NAD⁺ strongly supports enhanced NAD⁺ flux in $Nnmt$-ASO-treated adipocytes. In liver, $Nnmt$ knockdown did not alter NAD⁺ levels or the expression of $Nampt$, $Nmnat1$, $Nmnat2$ or $Nmnat3$ (Supplementary Fig. 20a, b). In spite of this, $Cd36$, Cat and $SdhB$ expression was decreased (Supplementary Fig. 20c), indicating reduced SIRT1 activity. Consistent with this, PGC-1α acetylation was enhanced in liver with $Nnmt$ knockdown[29] (Supplementary Fig. 20d). The decreased hepatic $Sirt1$ expression (Supplementary Fig. 20e) may contribute to the reduced SIRT1 activity with $Nnmt$ knockdown. Despite the changes in SIRT1 and PGC-1α activity, hepatic $Pck1$ and $G6pc$ expression were not altered by $Nnmt$ ASO treatment (Supplementary Fig. 21).

We discovered that NNMT is a novel regulator of adiposity and energy expenditure. This involves modulating adipose SAM and NAD⁺, two fundamental metabolites for energy metabolism. NNMT also regulates hepatic energy metabolism (Supplementary Fig. 9), but most likely with different mechanisms, as $Nnmt$ knockdown did not alter hepatic SAM and NAD⁺ levels. In adipocytes, SAM provides substrate for polyamine synthesis, and also modulates Odc and $Ssat$ expression by modifying H3K4 methylation (Fig. 4q). SAM availability may alter histone methylation in a methylation-site-specific manner rather than affecting global histone methylation[10]. For example, threonine-regulated SAM specifically affects H3K4me3 in mouse embryonic stem cells[30]. NNMT-regulated H3K4 methylation may have broader effects on gene expression including effects associated with cancer[10] and dementia, conditions in which NNMT activity is enhanced[10,11].

Activation of adipose polyamine flux causes leanness by catalysing polyamine acetylation to generate acetylpolyamines using acetyl-coA as a metabolic substrate[7]. Acetylpolyamines are oxidized in a futile cycle or excreted in the urine, thereby reducing acetyl-CoA in cells (Fig. 4q and Supplementary Fig. 12). Although urinary excretion of acetylpolyamines may not cause significant total-body calorie loss, the metabolic substrate consumption or loss in adipocytes seems to impact systemic

To determine whether the increased energy expenditure with NNMT inhibition (Fig. 3) is polyamine-dependent, we used N^1-methylnicotinamide to inhibit NNMT activity in adipocytes. This induced oxygen consumption (Fig. 4k–m). Knocking down $Ssat$ (Supplementary Fig. 17), inhibiting ODC activity[8] or blocking PAO activity abolished N^1-methylnicotinamide-induced oxygen consumption (Fig. 4k–m). Thus, NNMT inhibition in adipocytes autonomously enhances oxygen consumption, and this depends on polyamine flux.

The other substrate of the NNMT reaction is nicotinamide, an NAD⁺ precursor. $Nnmt$ ASO treatment did not alter adipose and hepatic nicotinamide levels (Supplementary Fig. 18a, b). This is likely to be because nicotinamide is metabolized to NAD⁺ in a salvage pathway, which is controlled by the rate-limiting enzymes nicotinamide phosphoribosyltransferase (NAMPT) and three isoforms of nicotinamide mononucleotide adenylyltransferase (NMNAT1, NMNAT2 and NMNAT3)[5]. $Nnmt$ knockdown increased NAD⁺ levels per mg of adipose tissue (Fig. 4n) and the expression of $Nampt$ (Fig. 4o), cytosolic $Nmnat2$ and mitochondrial $Nmnat3$, but not nuclear $Nmnat1$ (Supplementary Fig. 19). NAD⁺ is a cofactor of SIRT1 deacetylase activity, and SIRT1 affects energy metabolism[5]. Expression of SIRT1 target genes, including $Cd36$, catalase (Cat), succinate dehydrogenase B ($Sdhb$) and growth-arrest and DNA-damage-inducible protein, were increased in adipose tissue of $Nnmt$-ASO-treated mice[25–28] (Fig. 4p). As

energy expenditure. This is clearly demonstrated by leanness and increased energy expenditure in adipose *Ssat* transgenic mice[24]. Promoting consumption or loss of adipose acetyl-CoA as a metabolic substrate in the form of acetylpolyamines may be a novel strategy for obesity treatment (Fig. 4q).

NNMT regulates SAM and NAD^+ primarily in adipose tissue, but not in liver, which has redundant regulatory pathways for SAM and NAD^+ metabolism. Hepatic SAM is largely regulated by glycine *N*-methyltransferase[22] and by methionine, an abundant SAM precursor in liver. When methionine is abundant, NNMT regulates only SAH, not SAM[10]. Liver NAD^+ synthesis is controlled by a salvage pathway using nicotinamide as a precursor and a *de novo* pathway using tryptophan[5]. In contrast, adipose tissue barely expresses the key enzymes in the *de novo* NAD^+ synthesis pathway (data not shown). Therefore, adipose NAD^+ synthesis primarily relies on the salvage pathway using nicotinamide, for which NNMT is the only catabolic enzyme. The adipose-selective effects of NNMT inhibition on SAM and NAD^+ levels are advantageous for drug development because increased SAM in liver causes hepatic steatosis and fibrosis[22], and increased NAD^+ may enhance hepatic gluconeogenesis[29].

In summary, NNMT is a unique regulator of adiposity by directly altering NAD^+ and SAM, which affect histone methylation, polyamine flux and SIRT1 signalling. NNMT inhibition leads to metabolic substrate consumption or loss from adipocytes coupled with increased energy expenditure in a cell-autonomous manner (Fig. 4q). These unique features render NNMT an attractive target for treating obesity and type 2 diabetes.

METHODS SUMMARY

Two *Nnmt* ASOs were selected from 80 ASOs based on *Nnmt* knockdown efficiency and no hepatic or renal toxicity. Treatment of HFD-fed mice with either of the *Nnmt* ASOs resulted in leanness. The most potent *Nnmt* ASO was used for subsequent studies. An ASO (ISIS 425851) with no complementarity to any known gene sequence was used as the control. *Nnmt* ASOs and control ASOs were injected (37.5 mg kg^{-1} intraperitoneally) twice per week, starting 2 weeks after initiation of HFD feeding. ODC and SSAT activities were measured under initial rate conditions with non-limiting substrate concentrations as described[8]. Metabolites were measured using high-performance liquid chromatography (HPLC) with mass spectrometry (MS) using isotope dilution or targeted tandem mass spectrometry (LC–MS/MS)[30]. Diacetylspermine was measured using enzyme-linked immunosorbant assay (ELISA; Abnova). Two-tailed student's *t*-tests were used for single comparisons and analysis of variance (ANOVA) with Bonferroni-Holm post-hoc tests for multiple comparisons. Statistical significance is assumed if $P < 0.05$. Data are mean ± s.e.m.

1. Shepherd, P. R. & Kahn, B. B. Glucose transporters and insulin action—implications for insulin resistance and diabetes mellitus. *N. Engl. J. Med.* **341,** 248–257 (1999).
2. Yang, Q. *et al.* Serum retinol binding protein 4 contributes to insulin resistance in obesity and type 2 diabetes. *Nature* **436,** 356–362 (2005).
3. Aksoy, S., Szumlanski, C. L. & Weinshilboum, R. M. Human liver nicotinamide *N*-methyltransferase. cDNA cloning, expression, and biochemical characterization. *J. Biol. Chem.* **269,** 14835–14840 (1994).
4. Riederer, M., Erwa, W., Zimmermann, R., Frank, S. & Zechner, R. Adipose tissue as a source of nicotinamide *N*-methyltransferase and homocysteine. *Atherosclerosis* **204,** 412–417 (2009).
5. Houtkooper, R. H., Canto, C., Wanders, R. J. & Auwerx, J. The secret life of NAD^+: an old metabolite controlling new metabolic signaling pathways. *Endocr. Rev.* **31,** 194–223 (2010).
6. Teperino, R., Schoonjans, K. & Auwerx, J. Histone methyl transferases and demethylases; can they link metabolism and transcription? *Cell Metab.* **12,** 321–327 (2010).
7. Jell, J. *et al.* Genetically altered expression of spermidine/spermine N1-acetyltransferase affects fat metabolism in mice via acetyl-CoA. *J. Biol. Chem.* **282,** 8404–8413 (2007).
8. Pirinen, E. *et al.* Enhanced polyamine catabolism alters homeostatic control of white adipose tissue mass, energy expenditure, and glucose metabolism. *Mol. Cell. Biol.* **27,** 4953–4967 (2007).
9. Sartini, D. *et al.* Nicotinamide *N*-methyltransferase in non-small cell lung cancer: promising results for targeted anti-cancer therapy. *Cell Biochem. Biophys.* **67,** 865–873 (2013).
10. Ulanovskaya, O. A., Zuhl, A. M. & Cravatt, B. F. NNMT promotes epigenetic remodeling in cancer by creating a metabolic methylation sink. *Nature Chem. Biol.* **9,** 300–306 (2013).
11. Williams, A. C., Cartwright, L. S. & Ramsden, D. B. Parkinson's disease: the first common neurological disease due to auto-intoxication? *QJM* **98,** 215–226 (2005).
12. Lee, Y. H. *et al.* Microarray profiling of isolated abdominal subcutaneous adipocytes from obese vs non-obese Pima Indians: increased expression of inflammation-related genes. *Diabetologia* **48,** 1776–1783 (2005).
13. Salek, R. M. *et al.* A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human. *Physiol. Genomics* **29,** 99–108 (2007).
14. Yaguchi, H., Togawa, K., Moritani, M. & Itakura, M. Identification of candidate genes in the type 2 diabetes modifier locus using expression QTL. *Genomics* **85,** 591–599 (2005).
15. Wu, C. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* **10,** R130 (2009).
16. Alexander, J., Chang, G. Q., Dourmashkin, J. T. & Leibowitz, S. F. Distinct phenotypes of obesity-prone AKR/J, DBA2J and C57BL/6J mice compared to control strains. *Int. J. Obes. (Lond.)* **30,** 50–59 (2006).
17. Svenson, K. L. *et al.* Multiple trait measurements in 43 inbred mouse strains capture the phenotypic diversity characteristic of human populations. *J. Appl. Physiol.* **102,** 2369–2378 (2007).
18. Grubb, S. C., Maddatu, T. P., Bult, C. J. & Bogue, M. A. Mouse phenome database. *Nucleic Acids Res.* **37,** D720–D730 (2009).
19. Bennett, C. F. & Swayze, E. E. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu. Rev. Pharmacol. Toxicol.* **50,** 259–293 (2010).
20. Erion, D. M. *et al.* SirT1 knockdown in liver decreases basal hepatic glucose production and increases hepatic insulin responsiveness in diabetic rats. *Proc. Natl Acad. Sci. USA* **106,** 11288–11293 (2009).
21. Kang-Lee, Y. A. *et al.* Metabolic effects of nicotinamide administration in rats. *J. Nutr.* **113,** 215–221 (1983).
22. Varela-Rey, M. *et al.* Fatty liver and fibrosis in glycine *N*-methyltransferase knockout mice is prevented by nicotinamide. *Hepatology* **52,** 105–114 (2010).
23. Pegg, A. E. & Casero, R. A., Jr. Current status of the polyamine research field. *Methods Mol. Biol.* **720,** 3–35 (2011).
24. Koponen, T. *et al.* The activation of hepatic and muscle polyamine catabolism improves glucose homeostasis. *Amino Acids* **42,** 427–440 (2011).
25. Alcendor, R. R. *et al.* Sirt1 regulates aging and resistance to oxidative stress in the heart. *Circ. Res.* **100,** 1512–1521 (2007).
26. Finley, L. W. *et al.* Succinate dehydrogenase is a direct target of sirtuin 3 deacetylase activity. *PLoS ONE* **6,** e23295 (2011).
27. Kobayashi, Y. *et al.* SIRT1 is critical regulator of FOXO-mediated transcription in response to oxidative stress. *Int. J. Mol. Med.* **16,** 237–243 (2005).
28. Stein, S. *et al.* SIRT1 decreases Lox-1-mediated foam cell formation in atherogenesis. *Eur. Heart J.* **31,** 2301–2309 (2010).
29. Rodgers, J. T. *et al.* Nutrient control of glucose homeostasis through a complex of PGC-1α and SIRT1. *Nature* **434,** 113–118 (2005).
30. Shyh-Chang, N. *et al.* Influence of threonine metabolism on *S*-adenosylmethionine and histone methylation. *Science* **339,** 222–226 (2013).

METHODS

Mice. Adipose-specific *Glut4*-knockout and *Glut4*-overexpressing mice were described previously[31,32]. All other mice were purchased from Jackson Laboratories or Charles River Laboratories. For measurements of *Nnmt* expression in adipose tissue and liver, the *ob/ob* and *db/db* mice were killed at 10 to 12 weeks of age. High-fat diet (HFD)-fed male mice in C57BL/6 background were treated with a HFD for 12 weeks starting at 6 to 8 weeks of age. For *Nnmt* ASO experiments, mice were fed standard chow providing 17% calories from fat (LabDiet Formulab 5008) or a HFD providing 55% calories from fat (Harlan Teklad, TD93075) starting at 6 to 7 weeks of age. *Nnmt* and control ASOs were injected at a dose of 37.5 mg kg^{-1} intraperitoneally twice per week starting 2 weeks after the initiation of HFD feeding. Mouse studies were conducted in accordance with federal guidelines and were approved by the Beth Israel Deaconess Medical Centre (BIDMC) Institutional Animal Care and Use Committee.

Selection of mouse *Nnmt* ASO. Rapid-throughput screens were performed in primary hepatocytes to identify mouse *Nnmt* antisense oligonucleotides. In brief, 80 ASOs were designed to the *Nnmt* mRNA sequence and tested in primary hepatocytes to identify *Nnmt* antisense oligonucleotides. *Nnmt* gene knockdown was screened by quantitative PCR. Eight ASOs were selected and further characterized in a dose–response screen *in vivo*. Two *Nnmt* ASOs with greatest effects on *Nnmt* knockdown in liver and adipose tissue with no toxicity to liver and kidney were used to treat mice on a HFD. Treatment with either of the *Nnmt* ASOs resulted in leanness compared with a control-ASO-treated group. The most potent *Nnmt* ASO (GAAATGAACCAGCAGGCCTT) was chosen for subsequent studies. A control ASO (ISIS 425851), which has no complementarity to any known gene sequence, was used for the control group. *Nnmt* and control ASOs have a uniform phosphorothioate backbone and a 20-base chimaeric design with a 2'-O-(methoxy)-ethyl (2'-MOE) modification on the first five and the last five bases. This modification enhances their binding affinity to complementary sequences and their resistance to the action of nucleases. ASOs may downregulate their target genes through an RNase H-dependent cleavage mechanism or a non-RNA-degrading mechanism by blocking translation of mRNA. Two pathways may be involved ASO entry into cells: first, a productive pathway using a vesicular transport system; and second, a non-productive pathway that results in accumulation of ASOs in lysosomal structures[19].

CLAMS. Energy expenditure was evaluated using a Comprehensive Lab Animal Monitoring System (Columbia Instruments)[33]. HFD-fed mice were treated with *Nnmt*- or control-ASO for 3 weeks. Body weights were not significantly different when the mice were subjected to CLAMS study. The mice were acclimated in the metabolic chambers for 2 days before the experiments. CO_2 and O_2 levels were collected every 32 min for each mouse during a period of 3 days. Food intake and activity were measured at regular intervals.

RNA extraction and quantitative PCR. RNA extraction and Taqman quantitative PCR were performed as described previously[2]. All Taqman primers were purchased from Applied Biosytems, except for the probe for adenosylmethionine decarboxylase (*Amd1*), which was designed using IDT software: forward 5'-GAG AGTGGAATTCGTGACCTG-3'; probe 5'-ACTGTTCAATCCTTGTGGCTAC TCGATG-3'; reverse 5'-TTCTGGTTCTGGAGTGATGTG-3'.

Western blot analysis and immunoprecipitation. NNMT was detected using a chicken anti-NNMT antibody from GenWay Biotech or the rabbit anti-NNMT antibody kindly provided by R. Weinshilboum[34]. Goat anti-UCP1 (M-17) and rabbit anti-PGC1α (H-300) were from Santa Cruz Biotechnology. Rabbit anti-acetylated lysine was from Cell Signaling Technology. For detecting PGC-1α acetylation, PGC-1α was immunoprecipitated from liver or cultured-cell lysates with anti-PGC-1α. The precipitated protein was subjected to western blot analysis using anti-acetylated lysine.

Body composition, body temperature, and faecal lipid excretion. Body composition was analysed in mice treated with *Nnmt* and control ASO for 5 weeks using an EchoMRI 3-in-1 instrument (Echo Medical Systems). To measure body composition after fasting, food was removed from *Nnmt*-ASO- or control-ASO-treated mice for 16 h. Body temperature was measured intractrectally using a Thermalert TH-5 thermometer (Physitemp). To analyse faecal lipid excretion, lipid content of faeces was extracted using chloroform:methanol (2:1) and air-dried under a fume hood.

Histology and determination of cell size. Tissues were fixed in 10% buffered formalin and subjected to H&E staining. To measure adipocyte cross-sectional area, four fields of vision of H&E-stained adipose tissue were digitally photographed under a microscope. Images were analysed using the software ImageJ (National Institutes of Health) using a previously published technique[35]. Images were converted to 16-bit grayscale and inverted; a threshold was applied to identify cell boundaries, and the cells were analysed with the 'Analyze Particles' command to obtain adipocyte area. Adipocyte volume was calculated from the radius using the formula $v = 4/3\pi r^3$.

Intraperitoneal glucose tolerance test. The food was removed from mice for 5 h. Intraperitoneal glucose tolerance test (IPGTT) was performed by intraperitoneal injection of 1 mg glucose per g of body weight. Glucose levels were measured at the indicated times.

Feed efficiency. Total calorie intake was computed from the measured food intake during the second and third weeks of ASO treatment. Feed efficiency was expressed as the body weight gain divided by the total calorie intake between the start and the end of the food intake measurement[36].

Urinary diacetylspermine measurement. Spot urine was collected from *Nnmt*-ASO- and control-ASO treated mice. Urinary diacetylspermine levels were measured using an ELISA kit from Abnova. An enzymatic mouse creatinine assay (Crystal Chem) was used for measuring creatinine levels. Because the mice drank water ad libitum, the urinary diacetylspermine levels were corrected by creatinine levels. Urinary creatinine levels were not different between *Nnmt*-ASO- and control-ASO-treated mice.

ODC and SSAT activity. ODC and SSAT activities were assayed from tissue supernatant fractions under initial rate conditions with non-limiting substrate concentrations as described previously[37,38]. In brief, ODC was assayed by measuring the release of $^{14}CO_2$ in a reaction with DL-[1-^{14}C]ornithine as the substrate. SSAT was assayed by measuring the formation of acetylated spermidine in a reaction with ^{14}C-CoA and spermidine as co-substrates.

NAD$^+$ measurement. NAD$^+$ levels were determined according to instructions provided with the NAD/NADH assay kit purchased from Abcam.

Histone methylation and ChIP-qPCR. Histone protein was extracted from adipose tissue treated with *Nnmt* ASO and control ASO using the EpiQuik Total Histone Extraction Kit from Epigentek. Eight types of histone methylation were measured by western blot analysis using the following antibodies: H3K4me2, H3K9me2, H3K27me2, H3K36me2 and H3K79me2 from Cell Signaling; and H3K4me1, H3K4me3 and H3K9me3 from Abcam. The levels of histone methylation were normalized to total H3 expression. For ChIP, chromatin was first extracted from 3T3-L1 adipocytes treated with or without 10 mM *N*-methylnicotinamide for 48 h. Immunoprecipitation was performed using EpiQuik Methyl-Histone H3-K4 ChIP kit from Epigentek. Enrichment of methylated H3K4 on *Odc* and *Ssat* genes was measured by real-time PCR using EpiTect ChIP qPCR primers from Qiagen (*Odc*, GPM1030188(+)01A; *Ssat*, GPM1055920(+)01A). An open reading frame free region (*Igx1a*) was used as a negative control for ChIP-qPCR. Control IgG showed minimum background among all the regions analysed.

Targeted mass spectrometry. Metabolites nicotinamide, *S*-adenosylmethionine (SAM), *S*-adenosylhomocysteine (SAH), AMP and ATP were measured using tandem mass spectrometry[30,39]. Metabolite extracts using 80% methanol (−80 °C) were dried by nitrogen. Samples were re-suspended using 20 μl LC/MS grade water, of which 10 μl were injected and analysed using a 5500 QTRAP triple quadrupole mass spectrometer (AB/Sciex) coupled to a Prominence HPLC system (Shimadzu) via selected reaction monitoring (SRM). The dwell time was 4 ms per SRM transition and the total cycle time was 1.89 s. Approximately 8 to 11 data points were acquired per detected metabolite. Samples were delivered to the MS using a 4.6 mm internal diameter × 10 cm Amide XBridge HILIC column (Waters) at 300 μl min^{-1}. Gradients were run starting from 85% buffer B (HPLC grade acetonitrile) to 35% B from 0 to 3.5 min; 35% B to 2% B from 3.5 to 11.5 min; 2% B was held from 11.5 to 16.5 min; 2% B to 85% B from 16.5 to 17.5 min; 85% B was held for 7 min to re-equilibrate the column. Buffer A consisted of 20 mM ammonium hydroxide and 20 mM ammonium acetate (pH 9.0) in 95:5 water:acetonitrile. Peak areas from the total ion current for each metabolite SRM transition were integrated using MultiQuant v2.0 software (AB/Sciex).

Cell culture. 3T3-L1 cells were differentiated to adipocytes using a standardized protocol[40]. Five to six days after differentiation, cells were transfected using Amaxa electroporation[41]. Hepatoma H2.35 cells were transduced with NNMT or β-galactosidase adenovirus for 24 h, and oxygen consumption was measured as described below.

Oxygen consumption. Oxygen consumption was measured with a Clark-type oxygen electrode (Rank Brothers). Cultured cells were trypsinized, counted in a hemocytometer, spun down, and resuspended in respiration buffer (PBS, 2% BSA, 4.5 g l-1 glucose, 120 mg l^{-1} sodium pyruvate) to 1×10^6 ml^{-1}. Partial oxygen pressure was recorded using a Rank Brothers digital model 10 controller and a PowerLab 4/30 data acquisition system (ADInstruments)[42].

Statistical test. All data are expressed as mean mean ± s.e.m. Two-tailed Student's *t*-tests were used for single comparisons. Analyses of variance were performed followed by Bonferroni-Holm post-hoc tests for multiple comparisons. Statistical significance is assumed at $P < 0.05$. Sample size was chosen based on results from pilot studies and our extensive experience in investigating metabolic physiology in mice. The sample exclusion criteria were determined before experiments for technical failures such as mis-injection of glucose in glucose tolerance test. For randomization, experimental groups of mice were stratified according to body weight

using an algorithm based on Kullback–Leibler divergences[43]. The CLAMS study, metabolic analyses and oxygen consumption in adipocytes were carried out in a blinded fashion, others were not.

31. Abel, E. D. *et al.* Adipose-selective targeting of the *GLUT4* gene impairs insulin action in muscle and liver. *Nature* **409,** 729–733 (2001).
32. Shepherd, P. R. *et al.* Adipose cell hyperplasia and enhanced glucose disposal in transgenic mice overexpressing GLUT4 selectively in adipose tissue. *J. Biol. Chem.* **268,** 22243–22246 (1993).
33. Bubolz, A. H. *et al.* Activation of endothelial TRPV4 channels mediates flow-induced dilation in human coronary arterioles: role of Ca^{2+} entry and mitochondrial ROS signaling. *Am. J. Physiol. Heart Circ. Physiol.* **302,** H634–H642 (2012).
34. Yan, L., Otterness, D. M., Craddock, T. L. & Weinshilboum, R. M. Mouse liver nicotinamide *N*-methyltransferase: cDNA cloning, expression, and nucleotide sequence polymorphisms. *Biochem. Pharmacol.* **54,** 1139–1149 (1997).
35. Chen, H. C. & Farese, R. V., Jr. Determination of adipocyte size by computer image analysis. *J. Lipid Res.* **43,** 986–989 (2002).
36. Bence, K. K. *et al.* Neuronal PTP1B regulates body weight, adiposity and leptin action. *Nature Med.* **12,** 917–924 (2006).
37. Bernacki, R. J. *et al.* Preclinical antitumor efficacy of the polyamine analogue N1, N11-diethylnorspermine administered by multiple injection or continuous infusion. *Clin. Cancer Res.* **1,** 847–857 (1995).
38. Jänne, J. & Williams-Ashman, H. G. On the purification of L-ornithine decarboxylase from rat prostate and effects of thiol compounds on the enzyme. *J. Biol. Chem.* **246,** 1725–1732 (1971).
39. Yang, X. *et al.* Using tandem mass spectrometry in targeted mode to identify activators of class IA PI3K in cancer. *Cancer Res.* **71,** 5965–5975 (2011).
40. Yan, Q. W. *et al.* The adipokine lipocalin 2 is regulated by obesity and promotes insulin resistance. *Diabetes* **56,** 2533–2540 (2007).
41. Eguchi, J. *et al.* Interferon regulatory factors are transcriptional regulators of adipogenesis. *Cell Metab.* **7,** 86–94 (2008).
42. Pulinilkunnil, T. *et al.* Adrenergic regulation of AMP-activated protein kinase in brown adipose tissue *in vivo*. *J. Biol. Chem.* **286,** 8798–8809 (2011).
43. Endo, A., Nagatani, F., Hamada, C. & Yoshimura, I. Minimization method for balancing continuous prognostic variables between treatment and control groups using Kullback–Leibler divergence. *Contemp. Clin. Trials* **27,** 420–431 (2006).

ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression

Hong Wen[1,2]*, Yuanyuan Li[3,4]*, Yuanxin Xi[5]*, Shiming Jiang[1], Sabrina Stratton[1], Danni Peng[1], Kaori Tanaka[1], Yongfeng Ren[3,4], Zheng Xia[5], Jun Wu[6], Bing Li[6], Michelle C. Barton[1,2,7], Wei Li[5], Haitao Li[3,4] & Xiaobing Shi[1,2,7]

Recognition of modified histones by 'reader' proteins plays a critical role in the regulation of chromatin[1]. H3K36 trimethylation (H3K36me3) is deposited onto the nucleosomes in the transcribed regions after RNA polymerase II elongation. In yeast, this mark in turn recruits epigenetic regulators to reset the chromatin to a relatively repressive state, thus suppressing cryptic transcription[2]. However, much less is known about the role of H3K36me3 in transcription regulation in mammals. This is further complicated by the transcription-coupled incorporation of the histone variant H3.3 in gene bodies[3]. Here we show that the candidate tumour suppressor ZMYND11 specifically recognizes H3K36me3 on H3.3 (H3.3K36me3) and regulates RNA polymerase II elongation. Structural studies show that in addition to the trimethyl-lysine binding by an aromatic cage within the PWWP domain, the H3.3-dependent recognition is mediated by the encapsulation of the H3.3-specific 'Ser 31' residue in a composite pocket formed by the tandem bromo–PWWP domains of ZMYND11. Chromatin immunoprecipitation followed by sequencing shows a genome-wide co-localization of ZMYND11 with H3K36me3 and H3.3 in gene bodies, and its occupancy requires the pre-deposition of H3.3K36me3. Although ZMYND11 is associated with highly expressed genes, it functions as an unconventional transcription co-repressor by modulating RNA polymerase II at the elongation stage. ZMYND11 is critical for the repression of a transcriptional program that is essential for tumour cell growth; low expression levels of *ZMYND11* in breast cancer patients correlate with worse prognosis. Consistently, overexpression of ZMYND11 suppresses cancer cell growth *in vitro* and tumour formation in mice. Together, this study identifies ZMYND11 as an H3.3-specific reader of H3K36me3 that links the histone-variant-mediated transcription elongation control to tumour suppression.

ZMYND11 (also known as BS69) is an adenovirus early region 1A (E1A)-interacting protein that functions as a co-repressor of E1A and cellular transcription factors including the c-Myb and ETS-2 oncoproteins[4–6]. It contains several histone 'reader' modules, including a plant homeodomain (PHD), a bromodomain and a PWWP domain (Fig. 1a), indicating a role in chromatin regulation. To determine whether these domains can recognize histone modification(s), we used a histone peptide array (Supplementary Table 1) and found that the recombinant PHD–bromo–PWWP (PBP) domains specifically bound H3K36me3 peptides (Fig. 1b), which was also confirmed by peptide pull-down assays (Fig. 1c). In addition, ZMYND11 PBP also bound H3K36me3 on bulk histones and the reconstituted H3K36-specific methyl-lysine analogue-containing nucleosomes (Extended Data Fig. 1a, b). Several PWWP domains can recognize methylation on histones H3K36 or H3K79 (refs 7–9); therefore we asked whether ZMYND11 PWWP alone was able to bind H3K36me3. The results showed that neither the isolated PWWP nor the PHD and bromodomain bound H3K36me3, whereas the tandem bromo–PWWP

retained a weak interaction (Fig. 1d and Extended Data Fig. 1c). Isothermal titration calorimetry (ITC) showed a dissociation constant (K_d) of 56 μM for PBP and 71 μM for bromo–PWWP to the H3K36me3 peptide, whereas weak or no binding to the di-, mono- and unmethylated H3K36 peptides (Fig. 1e and Extended Data Fig. 1d). Furthermore, deletion of the PWWP, bromo or PHD domains greatly diminished full-length ZMYND11 binding to H3K36me3 peptide and chromatin, whereas deletion of the MYND domain had no effect (Extended Data Fig. 1e, f).

To gain molecular insights into the reader function of ZMYND11, we solved the crystal structure of the tandem bromo–PWWP domain in its free state and in complex with H3K36me3 peptide at 1.95 Å and 2.0 Å, respectively (Extended Data Table 1 and Supplementary Discussion). Successful crystallization of the bromo–PWWP–H3K36me3 complex was facilitated by alanine mutation of D234 and E236 within the bromodomain to disrupt unfavourable packing contacts without affecting protein folding and histone binding (Extended Data Fig. 2a–e). The tandem bromo–PWWP domain adopts a V-shaped structure imparted by a 'kink' around a previously uncharacterized zinc finger (ZnF) motif (Fig. 1f, g). Several structural features imply that ZMYND11 bromodomain is unlikely to be a histone acetyl-lysine-binding module (Extended Data Fig. 2f–l). The PWWP domain adopts a five-bladed β-barrel fold with an extended carboxy (C)-terminal α-helix (α_P), with a conserved H3K36me3-binding aromatic cage[7–9] formed by F291 and W294 of the β1–β2 loop and F310 of the β3–β4 loop (Fig. 1h and Extended Data Fig. 3a–c). Alanine mutations abrogated ZMYND11 binding to the H3K36me3 peptide and chromatin without affecting the overall protein folding (Figs 1i, 2i and Extended Data Fig. 3d–f). Despite extensive interactions observed in crystals for the histone peptide, the relatively weak binding affinity suggests that an alternative binding mechanism is plausible. Indeed, the PWWP domain was previously identified as a DNA-binding motif[10]. Consistent with a recent report[11], we found that ZMYND11 bromo–PWWP bound to DNA at low micromolar level (Extended Data Fig. 3g). The higher binding affinity for DNA than that for histone peptides is in agreement with observations for some other PWWP domains[12,13]. Although not detected in our pull-downs using peptides and oligonucleotides, all these features suggest that a synergy between histone and DNA may contribute to ZMYND11 binding at the nucleosomal level (Supplementary Discussion).

The H3K36me3 peptides used in the biochemical and structural studies were histone H3.3-based peptides (Fig. 2a). The histone variant H3.3 possesses a signature motif 'S31 … A87AIG90' that is distinct from the 'A31 … S87AVM90' sequence signature of the canonical histone H3.1/2 (ref. 3). Our structure showed that the H3.3-specific 'S31' residue and its neighbouring T32 were encapsulated at the bottom of the 'bromo–ZnF–PWWP' valley (Fig. 2b), contributing to the ZMYND11–H3 interactions in addition to the binding of K36me3 by the PWWP
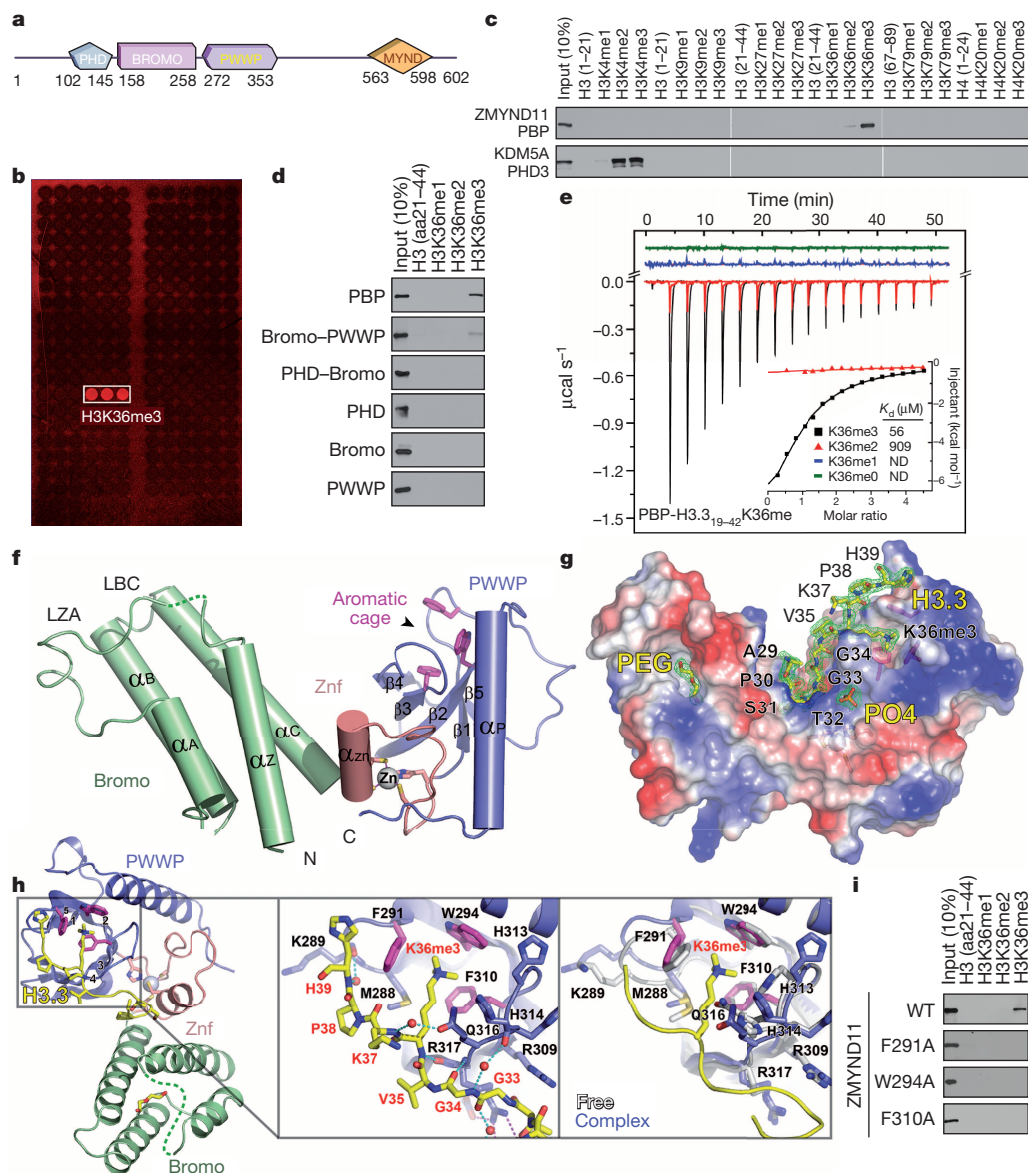
Figure 1 | Molecular basis for H3.3K36me3 recognition by ZMYND11 bromo–PWWP domains. a, Schematic representation of ZMYND11 protein structure. **b**, Histone peptide microarray (details in Supplementary Table 1) probed with glutathione S-transferase (GST)–ZMYND11 PBP domains. **c, d**, Western blot analysis of histone peptide pull-downs with GST–ZMYND11 PBP (**c**) or individual domains (**d**) and the indicated biotinylated peptides; aa, amino acids. **e**, ITC curves of histone H3.3K36me peptides titrated into ZMYND11 PBP domains; ND, not detectable. **f**, Cartoon view of ZMYND11 bromo–PWWP domains in the free state. **g**, Structure of ZMYND11 bromo–PWWP in complex with H3.3K36me3 peptide. Bromo–PWWP is shown as charge distribution surface. Blue, positive charges; red, negative charges. The

$2F_o − F_c$ omit maps around H3 peptide, polyethylene glycol (PEG) and phosphate (PO4) are shown as green mesh contoured at the 1σ level. **h**, Details of the K36me3 recognition by the PWWP domain. Close-up views: left, the hydrogen-bonding network around G34–H39; right, conformational adjustment of the PWWP aromatic cage upon H3K36me3 peptide binding. ZMYND11 bromo–PWWP in the free state is shown in light grey and bromo–PWWP in the complex state in dark blue; histone peptide is shown in yellow; small red balls indicate water. **i**, Western blot analysis of peptide pull-downs with GST–PBP or mutants and the indicated peptides. All the H3K36 peptides used in these experiments were H3.3-based.

aromatic cage. Indeed, the binding of ZMYND11 to H3.3K36me3 was much stronger than its binding to the H3.1 counterpart, with a seven- to eightfold difference in binding affinity (Fig. 2c, d and Extended Data Fig. 4a, b). Furthermore, Flag-tagged histone H3.3, but not H3.1, strongly co-immunoprecipitated ZMYND11 in cells (Extended Data Fig. 4c). Together, these results indicate that the H3.3-specific S31 residue is indispensable for ZMYND11–H3.3K36me3 interaction.

We also solved the structure of ZMYND11 bromo–PWWP–H3.1K36me3 complex at 2.3 Å for comparison (Extended Data Table 1 and Extended Data Fig. 4d). Several hydrogen-bonding interactions around the 'A29–G33' segment in the bromo–PWWP–H3.3 complex were lost in the bromo–PWWP–H3.1 complex after the 'Ser to Ala' conversion at position

31 (Fig. 2e, f and Extended data Fig. 4e, f). Superimposition of the complex structures showed a conformational adjustment of the H3.1 'A29–G33' segment (Fig. 2e), indicating a non-ideal encapsulation. As highlighted in the LIGPLOT diagram[14], eight direct hydrogen bonds and two sets of water-mediated hydrogen bonding were detected upon H3.3 variant binding (Fig. 2g), whereas only three direct hydrogen bonds and two sets of water-mediated hydrogen bonding were observed for H3.1 recognition (Fig. 2h). Notably, residues from all three modules, bromo, ZnF and PWWP, contributed to the A29–G33 segment recognition through the formation of a composite pocket, calling attention to the theme of 'module integration' for extended reader function[15]. Unique hydrogen-bonding interactions for the H3.3-specific readout were particularly
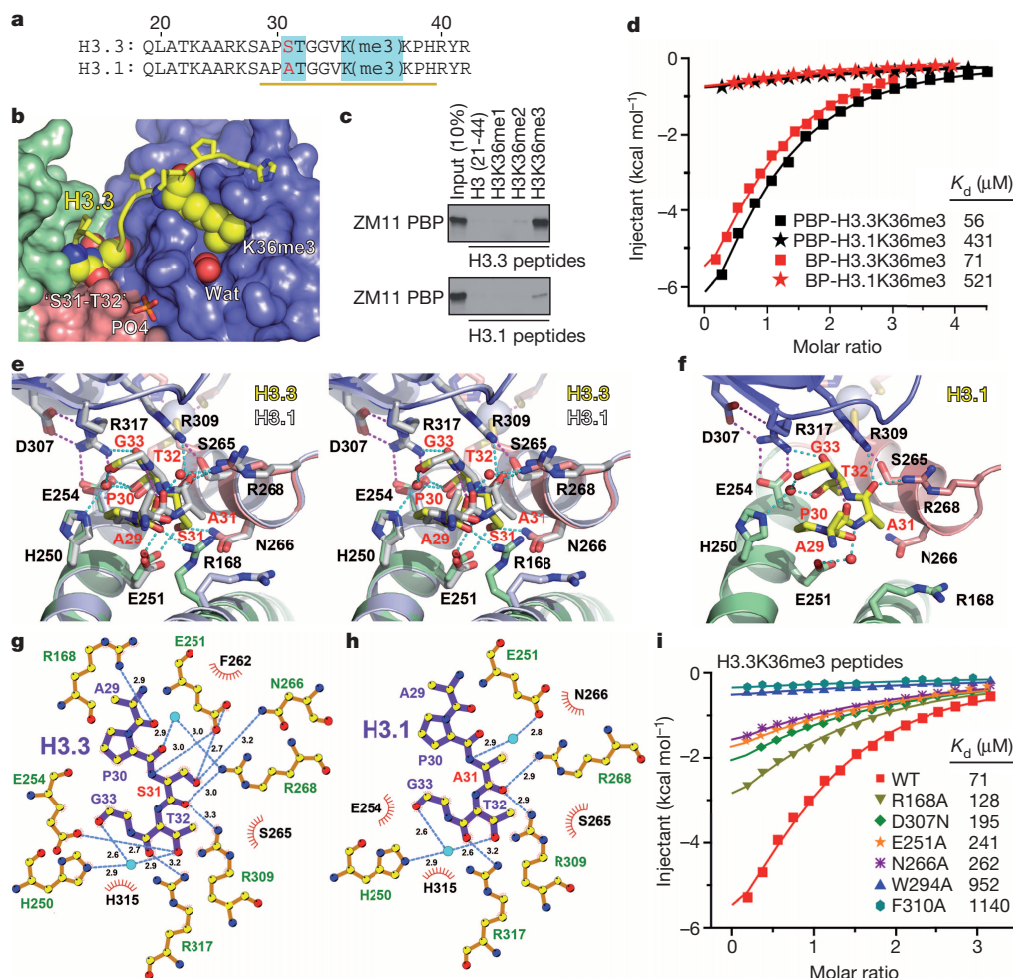
Figure 2 | H3.3 variant-specific K36me3 recognition by ZMYND11.
a, Sequence alignment of H3.3 and H3.1 spanning amino acids 19–42. The yellow line denotes the visible segment in the complex structures. **b,** Encapsulation of the 'S31–T32' segment and 'K36me3' by the paired bromo (pale green), ZnF (salmon) and PWWP (blue) modules of ZMYND11 rendered as solvent-accessible surface. The 'S31–T32' segment, K36me3 and two water (Wat) molecules are highlighted in spacing-filling mode. PO4, phosphate. **c,** Western blot analysis of peptide pull-downs with ZMYND11 PBP and the H3.3 or H3.1 variant peptides bearing K36me. **d,** ITC curves of H3.3K36me3 and H3.1K36me3 peptides titrated into ZMYND11 PBP or bromo–PWWP (BP) domains. **e, f,** Recognition of the 'A29–G33' segment of H3.3 (**e**) or H3.1 (**f**) by ZMYND11 at the junction of bromo (pale green), ZnF (salmon) and

PWWP (blue). Panel **e** is prepared in stereo view with H3.1 complex (grey) superimposed for comparison. Histone peptides are highlighted as yellow sticks. Hydrogen-bonding interactions are shown as dashes with cyan for those between H3 peptide and bromo–PWWP, and magenta for those within H3 peptide or bromo–PWWP. Small red balls, waters. **g, h,** LIGPLOT diagrams showing critical contacts between ZMYND11 bromo–PWWP and the 'A29-G33' segment of H3.3 (**g**) and H3.1 (**h**). H3 segment and key hydrogen-bonding residues from ZMYND11 bromo–PWWP are depicted in ball-and-stick mode. Cyan circle, water; blue dashed line, hydrogen bonds; curved brush, hydrophobic contact. **i,** ITC curves of ZMYND11 bromo–PWWP domains and the indicated point mutants with the H3.3K36me3 peptide.

supplied by 'R168–A29', 'E251–S31', 'E254–T32' and 'N266–S31' of the ZMYND11–H3.3 pair along with an adaptive flipping of the R168 side chain (Fig. 2e, g), and mutations diminished the bindings (Fig. 2i and Extended Data Fig. 4g). Together, the results strongly support a role of ZMYND11 as an H3.3-specific reader of H3K36me3.

We next sought to determine whether ZMYND11 co-localizes with H3K36me3 and the H3.3 variant in the genome by chromatin immuno-precipitation coupled with high-throughput sequencing (ChIP-seq) analysis. We identified 9,759 ZMYND11 peaks distributed in 3,353 genes in human osteosarcoma U2OS cells (Supplementary Table 2). The peaks were highly enriched in introns and exons (herein defined as the gene body) but not promoters (Fig. 3a). Strikingly, the average distribution of all ZMYND11-bound peaks recapitulated the genomic distribution of H3K36me3 (ref. 16) (Fig. 3b and Supplementary Table 3): absent from the transcription start site (TSS), but gradually increasing in the gene body towards the 3′ end and peaking at the transcription termination site (TTS).

We also determined the genome-wide distribution of histone variant H3.3 by ChIP-seq analysis of Flag–H3.3 stably expressed in U2OS cells

(Extended Data Fig. 5a and Supplementary Table 4). In accordance with previous observations[17–19], H3.3 was distributed in both the promoter and the body of active genes (Extended Data Fig. 5b). The gene-body-associated H3.3 showed a strong co-occupancy with H3K36me3 (Fig. 3c). Two-thirds of all the ZMYND11-bound genes overlapped with H3K36me3 and H3.3-co-occupied genes ($P < 1 \times 10^{-465}$) (Fig. 3c), as also represented by a genome-browser view of the ChIP-seq signals (Fig. 3d). High ZMYND11 occupancies were observed in genes enriched with both H3K36me3 and H3.3, whereas ZMYND11 occupancies were much lower in genes only decorated with H3K36me3 or H3.3 (Fig. 3e–g). Depletion of ZMYND11 by short hairpin RNAs (shRNAs) drastically diminished the ChIP-seq signals (Extended Data Fig. 5c–g), suggesting that the ChIP data obtained using our ZMYND11 antibody reflect the endogenous ZMYND11 chromatin-binding pattern.

ZMYND11 densities in gene bodies were directly correlated with H3K36me3 levels (Extended Data Fig. 6a). We then asked whether ZMYND11 occupancy was H3K36me3 dependent. We found that ZMYND11 binding on target genes was reduced along with diminished H3K36me3 levels in the SETD2 (ref. 20) knockdown cells (Fig. 3h and
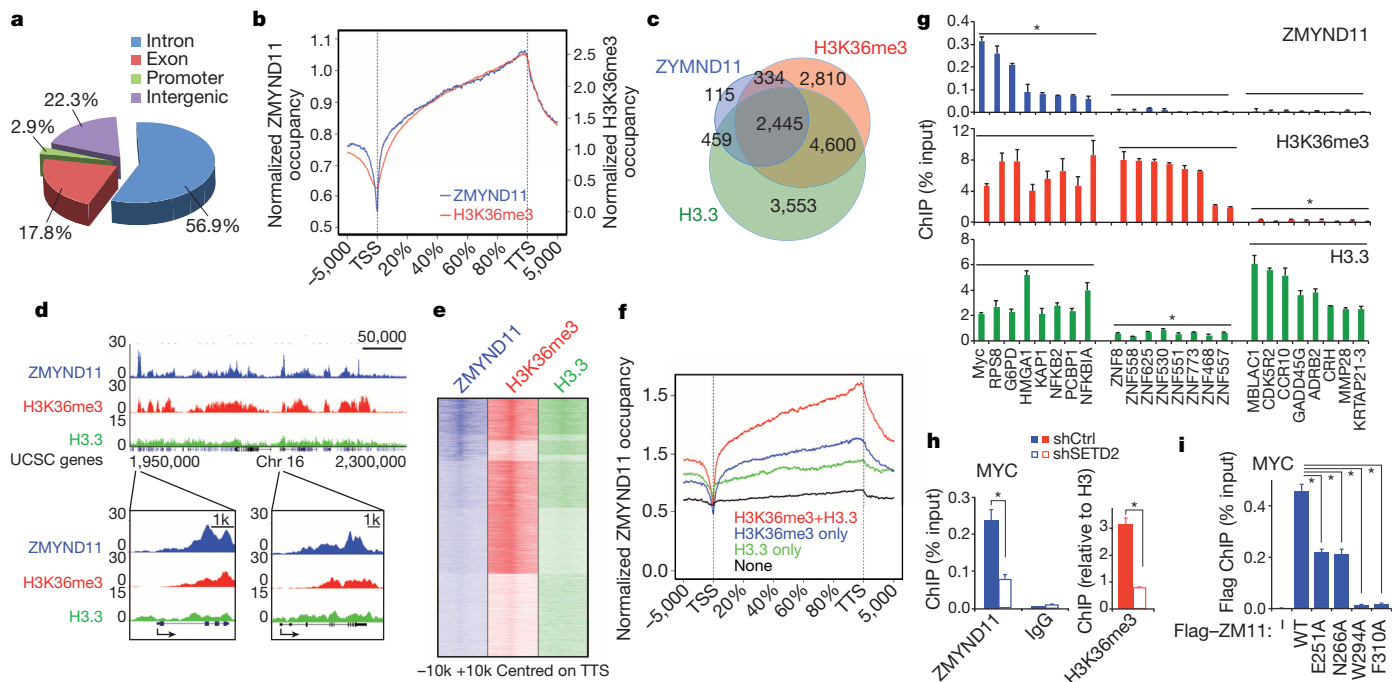
Figure 3 | ZMYND11 co-localizes with H3K36me3 and H3.3 in the gene body. a, Genomic distribution of ZMYND11 ChIP-seq peaks in U2OS cells. The peaks are enriched in introns and exons. $P < 1 \times 10^{-1208}$ (binomial test). **b,** Average genome-wide occupancies of ZMYND11 and H3K36me3 along the transcription unit. The gene body length is aligned by percentage from the TSS to TTS. Five kilobases upstream of TSS and 5 kb downstream of TTS are also included. **c,** Venn diagram showing the overlap of ZMYND11-, H3K36me3- and Flag–H3.3-occupied genes. $P < 1 \times 10^{-465}$ (three-way Fisher's exact test). **d,** Genome-browser view of ChIP-seq peaks in chromosome 16p13.3 regions. TSSs are indicated by arrows. **e,** Heatmaps of normalized density of ChIP-seq tags in a 20 kb window centred on TTS. **f,** Average genome-wide ZMYND11

occupancy on genes enriched with both H3K36me3 and H3.3 (red), genes with H3K36me3 (blue) or H3.3 (green) only, or genes with no H3K36me3 and H3.3 enrichment (black). Genes were aligned as described in **b**, and were grouped according to the H3K36me3 and Flag–H3.3 ChIP-seq normalized tag numbers >10 (high) or <2 (none). **g,** Quantitative PCR (qPCR) analysis of ZMYND11, H3K36me3 and Flag–H3.3 ChIP. **h,** qPCR analysis of ZMYND11 and H3K36me3 ChIP in *MYC* gene body in SETD2 knockdown cells. IgG ChIP is shown as a negative control. **i,** qPCR analysis of Flag ChIP in the cells stably expressing Flag-tagged wild type (WT) or mutant ZMYND11. In **g–i**, all error bars represent s.e.m. of three experiments. *$P < 0.05$.

Extended Data Fig. 6b–d), but not in the NSD2 (an H3K36me2 methyl-transferase)[21] knockdown cells where the H3K36me3 levels remained largely unchanged (Extended Data Fig. 6e, f). Furthermore, the H3.3K36me3 recognition-deficient mutants of ZMYND11 were severely impaired in binding on target genes (Fig. 3i and Extended Data Fig. 6g, h). Taken together, these results strongly suggest that ZMYND11 co-localizes with H3K36me3 and H3.3 in the gene body, and its occupancy requires the pre-deposition of trimethylation on H3.3K36.

To identify ZMYND11-regulated genes genome-wide, we performed RNA-seq analysis in U2OS cells. We found that 268 genes were upregulated and 370 genes downregulated upon ZMYND11 depletion (Fig. 4a, Extended Data Fig. 7a–c and Supplementary Table 5). Gene ontology analysis showed that the upregulated genes in ZMYND11 knockdown cells were specifically enriched in small cell lung cancer and focal adhesion pathways; in contrast, the downregulated genes were not enriched in any pathways (Fig. 4b and Supplementary Table 5). Ingenuity pathway analysis (IPA) of disease association of the upregulated genes also showed a high enrichment in cancer (Fig. 4b). These data suggest that ZMYND11 has a role in both transcriptional repression and activation, but it specifically represses oncogene expression.

Because ZMYND11 is localized in the gene body, we postulated that ZMYND11 regulates gene expression at the transcription elongation rather than the initiation step. The gene body RNA polymerase II (Pol II) density correlates well with transcription elongation rate[22,23]; we therefore determined the changes of Pol II density upon ZMYND11 depletion. ZMYND11 occupancies were enriched in genes with high Pol II densities ($P < 1 \times 10^{-401}$) (Extended Data Fig. 8a, b and Supplementary Table 6). Upon ZMYND11 depletion, we observed increased Pol II occupancy specifically on ZMYND11-repressed genes (Fig. 4c, d and

Extended Data Fig. 8c). Notably, only the amount of the intragenic Pol II, especially Pol II near the 3′ end of the gene body, was markedly increased, whereas the promoter-associated Pol II remained largely unchanged (Fig. 4c, d). Similar increases were also observed for the elongation-specific Pol II that is phosphorylated at serine 2 (S2P)[24] and the intragenic H3K36me3 levels (Fig. 4c and Extended Data Fig. 8d, e). The phenomenon of specific increase of Pol II in the gene body was further supported by the quantified Pol II travelling ratio[22] (Extended Data Fig. 8f). Together, these data suggest that ZMYND11 occupancy correlates with Pol II density in the gene body, and ZMYND11 represses gene expression by preventing the transition of paused Pol II to elongation.

ZMYND11 is downregulated in several human cancers including breast cancer (Extended Data Fig. 9a), indicating that ZMYND11 is a candidate tumour suppressor[25]. Overexpression of the wild-type ZMYND11 inhibited tumour cell growth, whereas the H3.3K36me3 binding-deficient mutants showed severe defects in suppressing cell proliferation and survival in both U2OS and the invasive MDA-MB 231 breast cancer cells (Fig. 4e and Extended Data Fig. 9b–e). Furthermore, in a mouse xenograft study using MDA-MB 231 cells, we found that compared with mice injected with control cells, tumour formation was strongly suppressed in mice injected with cells expressing wild-type ZMYND11, whereas the H3.3K36me3-binding deficient mutant W294A was severely impaired in suppressing tumour growth *in vivo* (Fig. 4f). Interestingly, D307N, a missense mutation identified in patients with colon and rectum adenocarcinoma (Extended Data Fig. 9f), diminished ZMYND11 binding to H3.3K36me3 (Fig. 2i and Extended Data Fig. 9g–i) and its tumour suppressor function in mice (Fig. 4f). Taken together, these results indicate that ZMYND11 suppresses phenotypes of cancer cells in a manner that depends on its H3.3K36me3-binding activity. Consistent
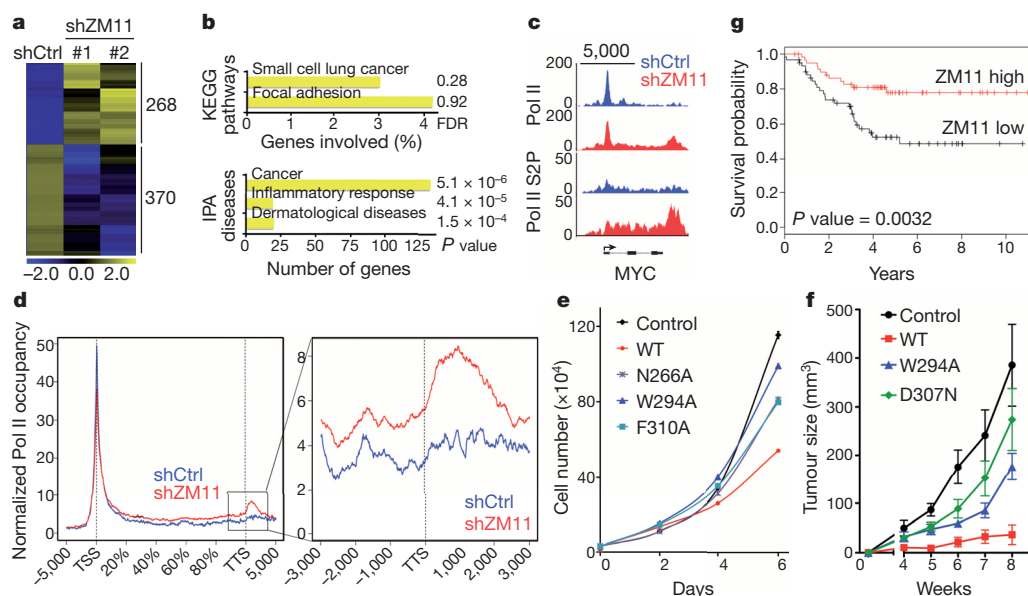
Figure 4 | ZMYND11 modulates Pol II elongation and represses oncogene expression and tumour growth. a, Heatmap representation of differentially expressed genes in control and ZMYND11 knockdown cells. Blue indicates relative low expression; yellow indicates high expression (details in Supplementary Table 5). **b,** KEGG and IPA pathway analyses of upregulated genes in ZMYND11 knockdown cells. Genes are shown as the percentage of genes (KEGG) or number of genes (IPA) within a functional group. False discovery rate (FDR) and P values are shown at the right. **c,** The genome browser view of Pol II and Pol II-S2P ChIP-seq signals in *MYC* gene in control and ZMYND11 knockdown cells. TSS is indicated with arrow. **d,** Average Pol II occupancy on ZMYND11-repressed direct target genes in control (blue) and ZMYND11 knockdown (red) cells. The left panel shows the entire gene

averages; the right panel shows the close-up view of the elongating Pol II near the TTS. $P < 2.47 \times 10^{-5}$ (two-way unpaired Student's *t*-test). **e,** Cell proliferation assay in MDA-MB 231 cells overexpressing wild-type ZMYND11 or the indicated mutants. Cells (mean ± s.e.m., $n = 4$) were counted for 6 days after seeding. **f,** Volumes of tumours (mean ± s.e.m., $n = 5$) derived from MDA-MB 231 cells overexpressing wild-type ZMYND11 or the indicated mutants over 8 weeks after subcutaneous xenograft transplants in immunodeficient nude mice. In **e, f,** $P < 0.001$ (one-way repeated-measures analysis of variance). **g,** Kaplan–Meier curves of overall recurrence-free survival of 120 patients with triple-negative molecular subtype of breast cancer[30]. Log-rank $P = 0.0032$ (χ^2 test).

with these experimental results, low ZMYND11 expression levels in breast cancer patients correlate with worse disease-free survivals (Fig. 4g and Extended Data 9j, k).

In summary, we have identified ZMYND11 as an H3.3-specific reader of H3K36me3. To our knowledge, it represents the first variant-specific reader of methylated histones. The combination of H3K36me3 and H3.3 establishes a unique epigenetic state that defines the genomic distribution of ZMYND11, offering a spatiotemporal control of gene expression for both normal and neoplastic growth. Because both H3K36me3 and H3.3 are deposited into the chromatin along with elongating Pol II, ZMYND11 is likely to be recruited to gene bodies only after one or several rounds of transcription when sufficient H3.3K36me3 is accumulated. We propose that, rather than as an on/off switch, ZMYND11 mainly functions in 'fine-tuning' gene expression by modulating Pol II elongation.

Our findings also emphasize the importance of the histone variant H3.3 in tumour suppression. Frequent mutations of the H3.3-encoding *H3F3A* gene have recently been identified in paediatric glioblastoma[26,27]. Unlike the K27M mutation in a single copy of the H3.3-encoding gene that affects the global H3K27 methylation amounts, the G34V/R mutations only alter the H3K36me3 levels at certain genes[28,29]. Interestingly, the G34V/R mutations impaired ZMYND11 binding to the H3.3K36me3 peptide (Extended data Fig. 9l). It remains an open question whether disruption of the binding of ZMYND11, or other yet unknown H3.3K36me3 readers, has a physiological impact on the pathogenesis of paediatric glioblastoma.

METHODS SUMMARY

Peptide microarray and binding assays were performed in buffer containing 50 mM Tris-HCl 7.5, 300 or 200 mM NaCl, 0.1% NP-40 overnight followed by three washes. Crystallizations of bromo–PWWP and bromo–PWWP–peptide complexes were performed by the hanging-drop vapour diffusion method at 18 °C.

Diffraction data were collected at synchrotron beamlines under cryoconditions. The initial phase was solved by zinc single-wavelength anomalous dispersion. Tumour xenograft was performed by subcutaneously injecting 5×10^6 MDA-MB 231 cells into nude mice. Survival analyses were performed as described previously[30]. Details are described in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Ruthenburg, A. J., Li, H., Patel, D. J. & Allis, C. D. Multivalent engagement of chromatin modifications by linked binding modules. *Nature Rev. Mol. Cell Biol.* **8,** 983–994 (2007).
2. Venkatesh, S. *et al.* Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* **489,** 452–455 (2012).
3. Elsaesser, S. J., Goldberg, A. D. & Allis, C. D. New functions for an old variant: no substitute for histone H3.3. *Curr. Opin. Genet. Dev.* **20,** 110–117 (2010).
4. Hateboer, G. *et al.* BS69, a novel adenovirus E1A-associated protein that inhibits E1A transactivation. *EMBO J.* **14,** 3159–3169 (1995).
5. Ladendorff, N. E., Wu, S. & Lipsick, J. S. BS69, an adenovirus E1A-associated protein, inhibits the transcriptional activity of c-Myb. *Oncogene* **20,** 125–132 (2001).
6. Wei, G., Schaffner, A. E., Baker, K. M., Mansky, K. C. & Ostrowski, M. C. Ets-2 interacts with co-repressor BS69 to repress target gene expression. *Anticancer Res.* **23,** 2173–2178 (2003).
7. Dhayalan, A. *et al.* The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* **285,** 26114–26120 (2010).
8. Vezzoli, A. *et al.* Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nature Struct. Mol. Biol.* **17,** 617–619 (2010).
9. Wu, H. *et al.* Structural and histone binding ability characterizations of human PWWP domains. *PLoS ONE* **6,** e18919 (2011).
10. Qiu, C., Sawada, K., Zhang, X. & Cheng, X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nature Struct. Biol.* **9,** 217–224 (2002).
11. Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152,** 1146–1159 (2013).
12. Eidahl, J. O. *et al.* Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Res.* **41,** 3924–3936 (2013).

13. van Nuland, R. *et al.* Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenet. Chromatin* **6,** 12 (2013).
14. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8,** 127–134 (1995).
15. Li, Y. & Li, H. Many keys to push: diversifying the 'readership' of plant homeodomain fingers. *Acta Biochim. Biophys. Sin.* **44,** 28–39 (2012).
16. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
17. Goldberg, A. D. *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140,** 678–691 (2010).
18. Jin, C. *et al.* H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nature Genet.* **41,** 941–945 (2009).
19. Ray-Gallet, D. *et al.* Dynamics of histone H3 deposition *in vivo* reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. *Mol. Cell* **44,** 928–941 (2011).
20. Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **27,** 406–420 (2008).
21. Kuo, A. J. *et al.* NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol. Cell* **44,** 609–620 (2011).
22. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141,** 432–445 (2010).
23. Danko, C. G. *et al.* Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol. Cell* **50,** 212–222 (2013).
24. Sims, R. J., III, Belotserkovskaya, R. & Reinberg, D. Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* **18,** 2437–2468 (2004).
25. Ansieau, S. & Sergeant, A. BS69 and RACK7, a potential novel class of tumor suppressor genes. *Pathol. Biol. (Paris)* **51,** 397–399 (2003).
26. Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482,** 226–231 (2012).
27. Sturm, D. *et al.* Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* **22,** 425–437 (2012).
28. Lewis, P. W. *et al.* Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* **340,** 857–861 (2013).
29. Bjerke, L. *et al.* Histone H3.3 mutations drive pediatric glioblastoma through upregulation of MYCN. *Cancer Discov.* **3,** 512–519 (2013).
30. Gyorffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res. Treat.* **123,** 725–731 (2010).

Author Contributions H.W., H.L. and X.S. conceived this study. H.W. performed the biochemical and cellular studies; Y.L. and H.L. performed structural and calorimetric studies; Y.X. and Z.X. performed bioinformatic analysis; S.J. and S.S. performed the xenograft study; D.P., K.T. and Y.R. provided technical assistance; J.W. and B.L. performed nucleosome binding experiments; X.S., H.L., H.W., W.L. and M.C.B. analysed data and wrote the paper.

Author Information Structure data have been deposited in Protein Data Bank under accession numbers 4N4G (free bromo–PWWP), 4N4H (bromo–PWWP–H3.1K36me3 complex) and 4N4I (bromo–PWWP–H3.3K36me3 complex). The ChIP-seq and RNA-seq data been deposited in the Gene Expression Omnibus database under accession number GSE48423. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.S. (xbshi@mdanderson.org), H.L. (lht@tsinghua.edu.cn), W.L. (WL1@bcm.edu), or H.W. (hwen@mdanderson.org).

LETTER

Direct measurement of local oxygen concentration in the bone marrow of live animals

Joel A. Spencer[1,2,3], Francesca Ferraro[4,5,6], Emmanuel Roussakis[1,7], Alyssa Klein[4,5,6], Juwell Wu[1,2], Judith M. Runnels[1,2], Walid Zaher[1,2,8], Luke J. Mortensen[1,2], Clemens Alt[1,2], Raphaël Turcotte[1,2,9], Rushdia Yusuf[4,5,6], Daniel Côté[10], Sergei A. Vinogradov[7], David T. Scadden[4,5,6] & Charles P. Lin[1,2,5]

Characterization of how the microenvironment, or niche, regulates stem cell activity is central to understanding stem cell biology and to developing strategies for the therapeutic manipulation of stem cells[1]. Low oxygen tension (hypoxia) is commonly thought to be a shared niche characteristic in maintaining quiescence in multiple stem cell types[2–4]. However, support for the existence of a hypoxic niche has largely come from indirect evidence such as proteomic analysis[5], expression of hypoxia inducible factor-1α (*Hif-1α*) and related genes[6], and staining with surrogate hypoxic markers (for example, pimonidazole)[6–8]. Here we perform direct *in vivo* measurements of local oxygen tension (p_{O_2}) in the bone marrow of live mice. Using two-photon phosphorescence lifetime microscopy, we determined the absolute p_{O_2} of the bone marrow to be quite low (<32 mm Hg) despite very high vascular density. We further uncovered heterogeneities in local p_{O_2}, with the lowest p_{O_2} (~9.9 mm Hg, or 1.3%) found in deeper peri-sinusoidal regions. The endosteal region, by contrast, is less hypoxic as it is perfused with small arteries that are often positive for the marker nestin. These p_{O_2} values change markedly after radiation and chemotherapy, pointing to the role of stress in altering the stem cell metabolic microenvironment.

The precise location of a haematopoietic stem cell (HSC) niche within the bone marrow (BM) remains elusive, with evidence supporting the existence of both a vascular niche[9,10] and an endosteal niche[11,12]. As the BM is densely perfused[13–16], including in the endosteal region, we ask how such a highly vascularized tissue can harbour HSCs in a low oxygen microenvironment. The possibility of very steep p_{O_2} gradients forming hypoxic zones at short distances away from the blood vessels is suggested by an *in silico* model[17,18]. However, no direct measurement of local oxygen distribution within the BM has been reported.

Here we implemented two-photon phosphorescence lifetime microscopy (2PLM) on a two-photon microscope designed specifically for live animal imaging[19]. The all-optical design (Extended Data Fig. 1) enabled non-contact p_{O_2} measurements through the intact skull at precise locations within the BM with micrometer spatial resolution. Before imaging, a metalloporphyrin-based two-photon-enhanced phosphorescent nanoprobe, platinum porphyrin-coumarin-343 (PtP-C343)[20,21], was injected systemically. The emissive triplet state of the PtP is highly sensitive to local oxygen concentration[20,21]. Bimolecular collisions with dissolved oxygen shorten the probe's triplet lifetime and quench phosphorescence[22]. Thus, by measuring the phosphorescence decay time τ after an excitation pulse, the absolute p_{O_2} value can be determined on the basis of a pre-established τ versus p_{O_2} calibration curve (Extended Data Fig. 2). The metalloporphyrin in PtP-C343 is protected by a dendrimer with a polyethylene glycol (PEG) overcoat (Extended Data Fig. 3) to insure biocompatibility and to prevent unwanted interactions of the probe with bio-macromolecules[20,23]. To enhance the two-photon

excitation cross-section, several C343 moieties are grafted onto the dendrimer shell as two-photon 'antennas' that funnel the captured energy to the metalloporphyrin by way of intramolecular energy transfer[20,21].

The residual fluorescence of the C343 units in the probe enabled visualization of the BM vasculature by conventional two-photon excited fluorescence, while the bone was visualized simultaneously by the collagen second harmonic generation (SHG) signal[14]. We confirmed that the BM contains high vascular density (Fig. 1a–c), with $>95\%$ of voxels in the image stack located <25 μm from the nearest blood vessel (Fig. 1b, c). We then quantified BM p_{O_2} at several intravascular locations using 2PLM and obtained values that ranged from 11.7 to 31.7 mm Hg (1.5–4.2%) with a mean of 20.4 mm Hg (2.7%). These values are significantly lower than the p_{O_2} in the microvasculature of the brain[24–26], the periosteum and the cortical bone (Fig. 1d). In many instances we were able to follow individual blood vessels as they penetrated from the bone into the BM cavity (Fig. 1e). When traced along individual vessels, we detected steep drops in p_{O_2}, measured immediately before and after entrance into the BM cavity (Fig. 1e, f). The rapid depletion of oxygen along the direction of blood flow, where the blood transits from a region of low cellularity (cortical bone) to a region of high cellularity (BM), is reminiscent of the longitudinal p_{O_2} gradient observed when blood vessels enter solid tumours with high metabolic demand[27].

Because of the relatively high permeability of the BM vasculature, the injected probe (diameter ~3–4 nm) (ref. 20) diffused out of the blood vessels and accumulated in sufficient concentrations within minutes (Extended Data Fig. 4), allowing p_{O_2} measurements in the interstitial space. In accordance with mathematical models[17,18], we found a steep lateral p_{O_2} gradient away from the blood vessels, with a mean extravascular p_{O_2} of 13.3 mm Hg (1.8%) and a range of 4.8 to 21.1 mm Hg (0.6–2.8%) (Fig. 1d).

Our measured p_{O_2} in the BM (~20.4 mm Hg intravascular, ~13.3 mm Hg extravascular) agreed well with a previously reported value of ~18 mm Hg obtained using an oxygen electrode[7]. However, that measurement lacked spatial resolution, and the insertion of the needle electrode probably damaged the microvasculature so that only an averaged reading of intravascular and extravascular p_{O_2} was recorded. Nevertheless, taken together these results support the notion that the BM as a whole is a hypoxic tissue despite its high vascularity.

To provide a finer-grained view of oxygen distribution within the BM, we measured p_{O_2} in different locations within the BM, as the vasculature is heterogeneous, with smaller vessels located closer to the endosteal surface[15] (Fig. 2a). 2PLM of blood vessels with diameters <15 μm shows higher p_{O_2} (22.7 mm Hg, 3.0%) than in blood vessels with diameters >15 μm (19.5 mm Hg, 2.6%; $P < 0.03$). Accordingly, when we analysed the p_{O_2} values at different distances from the endosteum, we found the lowest p_{O_2} to be in the regions >40 μm from the bone, with
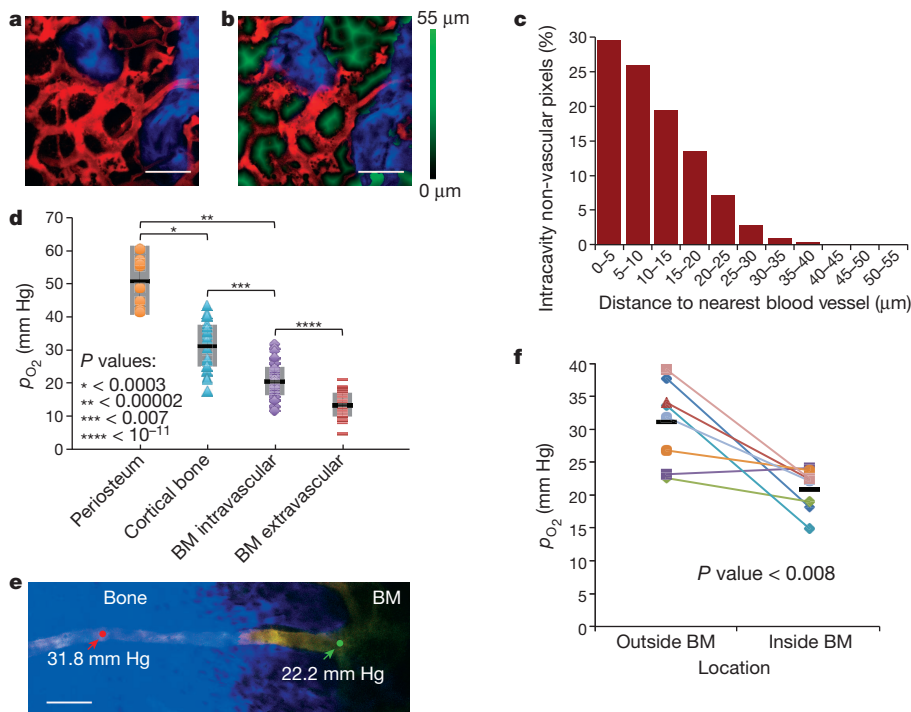
Figure 1 | BM vascular density and oxygenation.
a, Intravital maximum intensity image ($\sim$75 μm thick) of mouse calvarial BM showing blood vessels (red, Qtracker 655 vascular dye) and bone (blue, collagen SHG). **b**, Corresponding single plane of the three-dimensional intravital imaging stack showing blood vessels (red), bone SHG (blue), and the three-dimensional Euclidean distance measurement (EDM, green) to the nearest blood vessel wall for each extravascular pixel in the BM. **c**, Histogram of all EDMs from the full three-dimensional imaging stack. **d**, BM p_{O_2} is significantly lower compared with the p_{O_2} in periosteal and cortical bone vessels. Each point represents a p_{O_2} measurement in a separate blood vessel or interstitial position ($n = 8$, 21, 55 and 40 vessels/locations for periosteum, cortical bone, BM intravascular and BM extravascular, respectively, from 8 mice). The mean (black line) $\pm$ standard deviation (s.d.; shaded box) for each data set is shown. **e**, Maximum intensity projection image montage of a blood vessel entering the BM from the bone. Bone (blue) and blood vessels (yellow) are delineated with SHG and Rhodamine B–dextran/PtP-C343 fluorescence, respectively. The two arrows point to locations of p_{O_2} measurements just before and after the vessel enters the BM. **f**, Drop in p_{O_2} when tracking along individual vessels upon entry into the BM ($n = 8$ vessels from 4 mice). The mean (black line) is shown for each data set. Scale bars, $\sim$100 μm.

values of 17.7 mm Hg (2.4%) in the vessels and 9.9 mm Hg (1.3%) outside the vessels (Fig. 2a). In the endosteal region (0–20 μm zone), where most of the smaller vessels are located, the p_{O_2} readings were slightly higher, with values of 21.9 mm Hg (2.9%; $P < 0.03$) in the vessels and 13.5 mm Hg (1.8%; $P < 0.01$) outside the vessels (Fig. 2a). These measurements uncovered a moderate p_{O_2} gradient with distance from the endosteum, but the direction of the gradient was unexpected. Instead of the endosteal zone being the most hypoxic, we found that p_{O_2} decreased with increasing distance away from the endosteum towards the more hypoxic sinusoidal region.

As there are currently no HSC-specific markers for visualizing endogenous stem cells under homeostatic conditions, we used nestin-green fluorescent protein (nestin-GFP) mice to help identify the locations where HSCs reside. Nestin is expressed by a subset of BM mesenchymal stem cells that are involved in HSC maintenance and trafficking[9], and perivascular nestin-GFP bright cells define a quiescent HSC niche[16]. We reasoned that if hypoxia regulates quiescence[5], then nestin-GFP cells should be a landmark for hypoxic regions in the BM. In nestin-GFP mice, we confirmed that nestin[+] cells are predominantly perivascular, and transplanted HSCs home close to nestin[+] vessels[9] (Extended
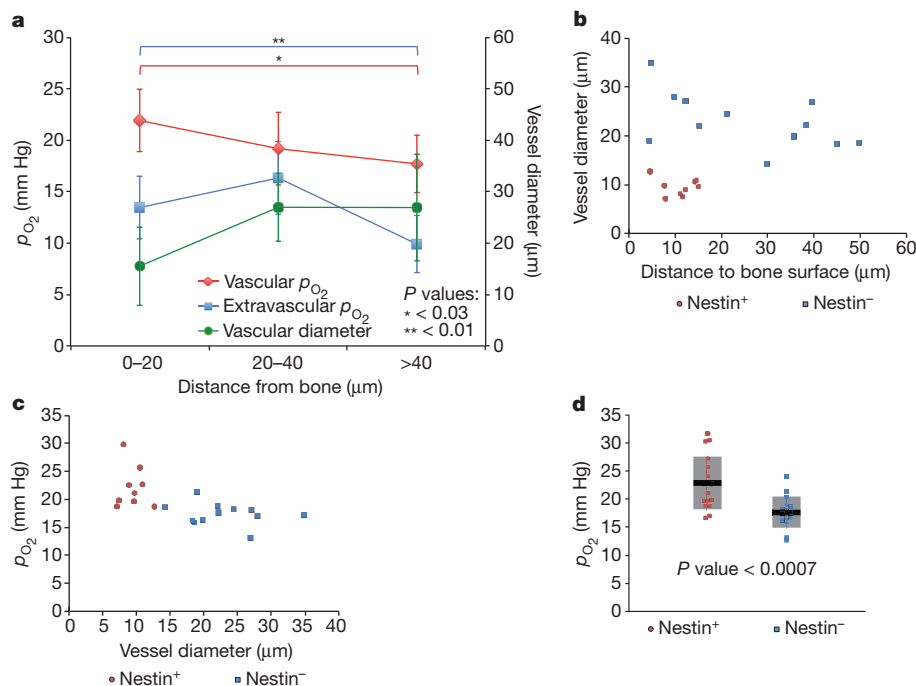

Figure 2 | Variation in BM p_{O_2} according to vessel diameter and distance from the endosteal surface. a, Average vessel diameter (green line; $n = 38$ vessels from 4 mice) and intravascular (red line; $n = 38$ vessels from 4 mice) and extravascular (blue line; $n = 39$ locations from 3 mice) p_{O_2} are plotted as a function of the distance from the bone surface to the measurement location. Error bars are $\pm$ s.d. **b**, The diameter of nestin[+] (red circles; $n = 9$ vessels from 3 mice) and nestin[−] vessels (blue squares; $n = 12$ vessels from 3 mice) are plotted as a function of the distance from the endosteal surface. **c**, The p_{O_2} of the same nestin[+] and nestin[−] vessels are plotted as a function of vessel diameter. **d**, Nestin[+] (red circles; $n = 17$ vessels from 7 mice) and nestin[−] (blue squares; $n = 16$ vessels from 5 mice) vessel p_{O_2}. The mean (black line) $\pm$ s.d. (shaded box) for each data set is shown.
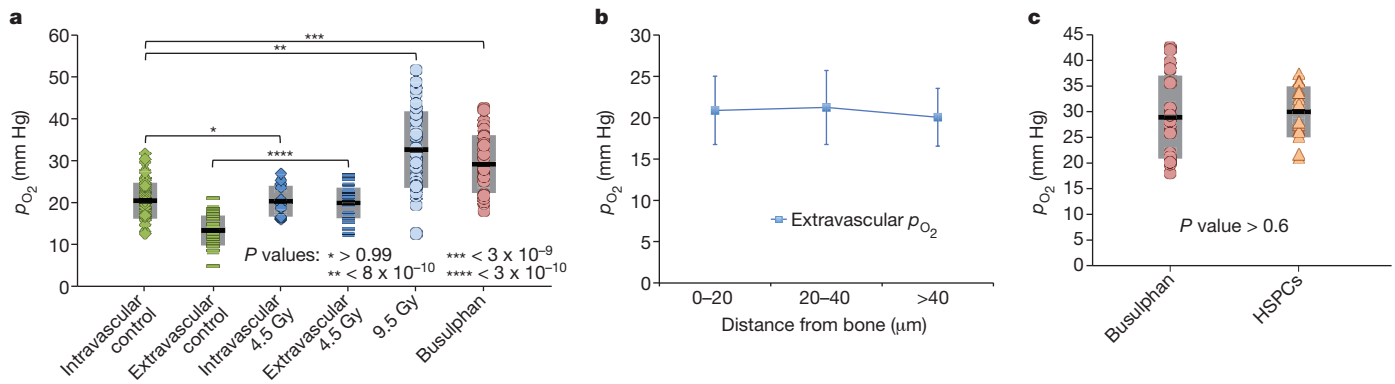
Figure 3 | BM p_{O_2} after cytoreductive therapy and at sites of HSPC homing.
a, BM p_{O_2} 2 days after sublethal (4.5 Gy; $n = 13$ vessels and 29 extravascular locations from 5 mice) or lethal (9.5 Gy; $n = 40$ locations from 2 mice) gamma irradiation, or after busulphan (35 mg kg^{-1}; $n = 40$ locations from 6 mice) conditioning. The untreated control p_{O_2} from Fig. 1d is plotted for comparison. **b**, Extravascular p_{O_2} measurements are plotted as a function of distance to the bone surface from the measurement location 2 days after 4.5 Gy irradiation ($n = 44$ locations from 4 mice). **c**, Comparison of p_{O_2} at sites of HSPC homing to the overall BM p_{O_2} in busulphan-treated recipients ($n = 17$ HSPCs from 6 mice). No statistically significant difference was observed. For **a** and **c**, the mean (black line) $\pm$ s.d. (shaded box) for each data set is shown.

Data Fig. 5). Nestin$^+$ vessels are anatomically distinct from nestin$^-$ vessels, being located closer to the bone surface (10.9 μm versus 25.5 μm; $P < 0.02$) and much smaller in diameter (9.5 μm versus 22.9 μm; $P < 2 \times 10^{-6}$) (Fig. 2b). Contrary to our expectation, we found p_{O_2} to be significantly higher in nestin$^+$ compared with nestin$^-$ vessels (22.8 mm Hg versus 17.6 mm Hg; $P < 0.0007$) (Fig. 2c, d).

Next we examined whether transplanted haematopoietic stem/progenitor cells (HSPCs) home to BM niches with distinct p_{O_2} levels. Before transplantation, recipient animals were given cytoreductive conditioning with either radiation or chemotherapy (busulphan, 35 mg kg^{-1}). Remarkably, both treatments resulted in a substantial elevation in the overall BM p_{O_2} (Fig. 3a). The rise in p_{O_2} was notable considering that the vasculature was severely damaged after both radiation and chemotherapy (Extended Data Fig. 6a–c). Moreover, although the intravascular p_{O_2} values were essentially unchanged after sub-lethal irradiation

(4.5 Gy), the gradient between intravascular and extravascular p_{O_2} disappeared, with the extravascular p_{O_2} rising to the same level as the intravascular p_{O_2} (Fig. 3a). The 'negative' gradient of decreasing p_{O_2} with increasing distance away from the endosteum also disappeared (Fig. 3b). In the case of lethal irradiation and busulphan treatment, the extensive vascular damage caused such widespread dye leakage that the boundary between intravascular and extravascular spaces could no longer be delineated.

We performed HSPC transplantation ($\sim 10^5$ lineage$^-$cKit$^+$Sca-1$^+$ (LKS) cells per recipient) 24 h after busulphan treatment and measured the p_{O_2} around each LKS cell that homed to the calvarial BM 18–24 h later. We obtained p_{O_2} values that span nearly the entire range of the (elevated) BM p_{O_2} in the post-chemotherapy setting (Fig. 3c), indicating that HSPCs did not seek out specific niches defined by low p_{O_2} as sites for preferential homing.
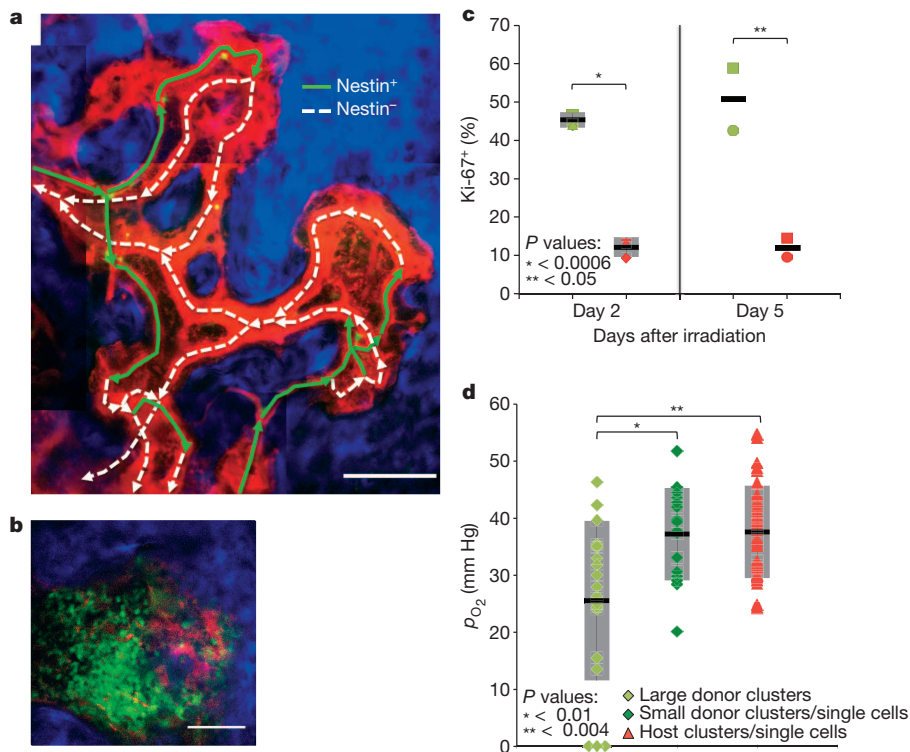

Figure 4 | BM vascular mapping and the effects of cellularity on local p_{O_2} after cytoreductive conditioning. **a**, Vascular connectivity map of a nestin-GFP mouse reveals that nestin$^+$ vessels are upstream of and drain into sinusoids. Arrows indicating blood flow direction, determined by tracking the movement of labelled RBCs with video-rate imaging, are superimposed onto a maximum intensity projection image of BM vasculature (red, vascular dye; green, nestin-GFP/RBCs; blue, SHG bone signal). A solid green line indicates a vessel with an adjacent nestin$^+$ cell and a dashed white line indicates a nestin$^-$ vessel. **b**, In vivo image of the bone marrow of a universal DsRed recipient 5 days after transplantation with 25 million total marrow cells from a universal GFP donor. Red, DsRed; green, GFP; blue, SHG from bone. **c**, Ki-67 staining of green (donor) and red (host) BM cells by FACS on day 2 ($n = 3$ mice) and day 5 ($n = 2$ mice) after transplantation. Each symbol denotes a different mouse. The mean (black line) for days 2 and 5 $\pm$ s.d. (shaded box) for day 2 is shown. **d**, The in vivo p_{O_2} values in regions with large clusters of donor cells ($n = 19$ locations from 4 mice) compared with small donor clusters/single cells ($n = 16$ locations from 4 mice) and host clusters/single cells ($n = 40$ locations from 5 mice). The mean (black line) $\pm$ s.d. (shaded box) for each data set is shown. Scale bars, ~ 100 μm.

These observations compelled us to re-examine blood flow in the BM[28]. Specifically we asked whether the differences in p_{O_2} observed in nestin[+] versus nestin[−] vessels and in steady-state versus radiation/chemotherapy-treated animals could be accounted for, at least in part, by the blood flow profiles. We injected carboxyfluorescein succinimidyl ester (CFSE)-labelled red blood cells (RBCs) and increased our imaging speed to 120 frames per second (scanning one quarter of the frame size) to enable the tracking of individual RBC flow direction and speed. Interestingly, the highest RBC flow velocities are observed in nestin[+] vessels (>2 mm s[−1], compared with 0.2–1 mm s[−1] in sinusoidal vessels). The vascular network map shows that the nestin[+] vessels are always upstream of and drain into sinusoidal vessels (Fig. 4a), suggesting that nestin[+] vessels have arterial characteristics. *In vivo* immunostaining[29] with fluorescent anti-Sca-1 antibody confirms that they are arteries[15] (Extended Data Fig. 7a).

Flow imaging with labelled RBCs also helps to overcome the dye leakage problem after radiation or chemotherapy and enables us to delineate the functional blood vessels by analysing the tracks of the flowing RBCs in the video sequence. We found a surprisingly robust flow 2–6 days after lethal irradiation (Extended Data Fig. 7b and Supplementary Video 1), a time when BM cellularity is declining precipitously (Extended Data Fig. 8). A combination of ample blood supply and reduced oxygen consumption (reduced BM cellularity) may explain the elevation of p_{O_2} after radiation and chemotherapy. Reduced BM cellularity may also explain why the gradient between intra- and extra-vascular p_{O_2} disappears after myelosuppressive conditioning.

To establish further the link between cellularity and oxygen consumption, we infused 25×10^6 total marrow cells from GFP donor mice into lethally irradiated DsRed recipients (Fig. 4b). We verified by fluorescence-activated cell sorting (FACS) on day 2 and 5 that most of the proliferating (Ki-67[+]) cells are in the donor fraction (Fig. 4c). Then, by *in vivo* imaging, we showed that the BM contained heterogeneous patches of donor (GFP[+]) and recipient (DsRed[+]) cells, with corresponding differences in local p_{O_2} (Fig. 4b). Lower p_{O_2} was associated with large clusters of green cells, consistent with the notion that the proliferating cells consume oxygen more avidly (Fig. 4d).

Direct measurements of the absolute p_{O_2} in the BM have revealed a unique hypoxic landscape organized by its dense vascularity (oxygen supply) and high cellularity (consumption). The balance between supply and consumption can be altered by stress such as radiation and chemotherapy. The local topography is further defined by the positioning of different types of blood vessels within the BM (Extended Data Fig. 9). In particular, HSCs near nestin[+] arteries and those near sinusoidal vessels will experience different metabolic milieus, highlighting the need to examine further the role of distinct vascular niches in HSC regulation.

METHODS SUMMARY

Intravital imaging and p_{O_2} measurement. All p_{O_2} measurements and intravital images were acquired with a home-built two-photon microscope (Extended Data Fig. 1). The arrival time of individual phosphorescence photons after pulse excitation was recorded using a custom photon counting circuit. The histogram of the photon arrival times was then analysed to obtain the triplet state lifetime.

For vascular mapping, RBCs were isolated from a donor C57BL/6 mouse, labelled with carboxyfluorescein diacetate succinimidyl ester (CFDA-SE; Life Technologies), and injected together with 200 μl of 400 nM Qtracker 655 (Life Technologies) or 200 μl of 3 mg ml[−1] 70 kDa Rhodamine B–dextran (Life Technologies) vascular label in PBS solution into the nestin-GFP mice immediately before imaging.

Image quantification. Using the axial stacks, the distance of the p_{O_2} measurement to the nearest bone surface (endosteum) was measured as described previously[14]. For distance measurements to blood vessels, three-dimensional stacks were resliced using Fiji (Image JA v.1.45b) into the vertical planes (*Y–Z, X–Z*) and digitally rotated to remove the tilt of the skull relative to the imaging plane. The three-dimensional stack was converted into binary images and manually verified for structural accuracy. We applied the exact Euclidean distance transform (EEDT) function on the binarized red (blood vessel) channel stack to determine the three-dimensional distance of each non-vascular pixel to the nearest blood vessel wall.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Lymperi, S., Ferraro, F. & Scadden, D. T. The HSC niche concept has turned 31. *Ann. NY Acad. Sci.* **1192,** 12–18 (2010).
2. Suda, T., Takubo, K. & Semenza, G. L. Metabolic regulation of hematopoietic stem cells in the hypoxic niche. *Cell Stem Cell* **9,** 298–310 (2011).
3. Mohyeldin, A., Garzón-Muvdi, T. & Quiñones-Hinojosa, A. Oxygen in stem cell biology: a critical component of the stem cell niche. *Cell Stem Cell* **7,** 150–161 (2010).
4. Lee, K. E. & Simon, M. C. From stem cells to cancer stem cells: HIF takes the stage. *Curr. Opin. Cell Biol.* **24,** 232–235 (2012).
5. Unwin, R. D. Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells. *Blood* **107,** 4687–4694 (2006).
6. Takubo, K., Goda, N., Yamada, W., Iriuchishima, H. & Ikeda, E. Regulation of the HIF-1α level is essential for hematopoietic stem cells. *Cell Stem Cell* **7,** 391–402 (2010).
7. Ceradini, D. J. *et al.* Progenitor cell trafficking is regulated by hypoxic gradients through HIF-1 induction of SDF-1. *Nature Med.* **10,** 858–864 (2004).
8. Parmar, K., Mauch, P., Vergilio, J.-A., Sackstein, R. & Down, J. D. Distribution of hematopoietic stem cells in the bone marrow according to regional hypoxia. *Proc. Natl Acad. Sci. USA* **104,** 5431–5436 (2007).
9. Méndez-Ferrer, S. *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466,** 829–834 (2010).
10. Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121,** 1109–1121 (2005).
11. Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425,** 841–846 (2003).
12. Wang, L. D. L. & Wagers, A. J. A. Dynamic niches in the origination and differentiation of haematopoietic stem cells. *Nature Rev. Mol. Cell Biol.* **12,** 643–655 (2011).
13. Lichtman, M. A. M. The ultrastructure of the hemopoietic environment of the marrow: a review. *Exp. Hematol.* **9,** 391–410 (1981).
14. Lo Celso, C. *et al.* Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche. *Nature* **457,** 92–96 (2009).
15. Nombela-Arrieta, C. *et al.* Quantitative imaging of haematopoietic stem and progenitor cell localization and hypoxic status in the bone marrow microenvironment. *Nature Cell Biol.* **15,** 533–543 (2013).
16. Kunisaki, Y. *et al.* Arteriolar niches maintain haematopoietic stem cell quiescence. *Nature* **502,** 637–643 (2013).
17. Chow, D. C., Wenning, L. A., Miller, W. M. & Papoutsakis, E. T. Modeling pO₂ distributions in the bone marrow hematopoietic compartment. I. Krogh's model. *Biophys. J.* **81,** 675–684 (2001).
18. Chow, D. C., Wenning, L. A., Miller, W. M. & Papoutsakis, E. T. Modeling pO₂ distributions in the bone marrow hematopoietic compartment. II. Modified Kroghian models. *Biophys. J.* **81,** 685–696 (2001).
19. Veilleux, I., Spencer, J. A., Biss, D. P., Cote, D. & Lin, C. P. *In vivo* cell tracking with video rate multimodality laser scanning microscopy. *IEEE J. Sel. Topics Quantum Electron.* **14,** 10–18 (2008).
20. Lebedev, A. Y., Troxler, T. & Vinogradov, S. A. Design of metalloporphyrin-based dendritic nanoprobes for two-photon microscopy of oxygen. *J. Porphyr. Phthalocyanines* **12,** 1261–1269 (2008).
21. Finikova, O. S. Oxygen microscopy by two-photon-excited phosphorescence. *ChemPhysChem* **9,** 1673–1679 (2008).
22. Vanderkooi, J. M. J., Maniara, G. G., Green, T. J. T. & Wilson, D. F. D. An optical method for measurement of dioxygen concentration based upon quenching of phosphorescence. *J. Biol. Chem.* **262,** 5476–5482 (1987).
23. Lebedev, A. Y. *et al.* Dendritic phosphorescent probes for oxygen imaging in biological systems. *ACS Appl. Mater. Interfaces* **1,** 1292–1304 (2009).
24. Sakadžić, S. *et al.* Two-photon high-resolution measurement of partial pressure of oxygen in cerebral vasculature and tissue. *Nature Methods* **7,** 755–759 (2010).
25. Lecoq, J. *et al.* Simultaneous two-photon imaging of oxygen and blood flow in deep cerebral vessels. *Nature Med.* **17,** 893–898 (2011).
26. Kazmi, S. M. S. *et al.* Three-dimensional mapping of oxygen tension in cortical arterioles before and after occlusion. *Biomed. Opt. Express* **4,** 1061–1073 (2013).
27. Dewhirst, M. W. M. *et al.* Quantification of longitudinal tissue pO₂ gradients in window chamber tumours: impact on tumour hypoxia. *Br. J. Cancer* **79,** 1717–1722 (1999).
28. Mazo, I. B. *et al.* Total body irradiation causes profound changes in endothelial traffic molecules for hematopoietic progenitor cell recruitment to bone marrow. *Blood* **99,** 4182–4191 (2002).
29. Sipkins, D. A. *et al.* *In vivo* imaging of specialized bone marrow endothelial microdomains for tumour engraftment. *Nature* **435,** 969–973 (2005).

Supplementary Information is available in the online version of the paper.

CORRECTIONS & AMENDMENTS

Corrigendum: Enhanced bacterial clearance and sepsis resistance in caspase-12-deficient mice

Maya Saleh, John C. Mathison, Melissa K. Wolinski, Steve J. Bensinger, Patrick Fitzgerald, Nathalie Droin, Richard J. Ulevitch, Douglas R. Green & Donald W. Nicholson

Owing to an error in the production process, some details were omitted from the advance online publication version of this Corrigendum: this is the complete version. When our Letter was under consideration at *Nature*, we originally showed co-immunoprecipitation between caspase-1 and wild-type caspase-12 or catalytically inactive caspase-12 (C299A) as part of Fig. 4. In response to reviewers' comments, requesting co-immunoprecipitation with other caspases for specificity control, this original figure was removed from the manuscript and was later published as part of figure 6 of ref. 1. It was recently brought to our attention that the published Fig. 4c of our Letter is a composite image containing parts of the original figure (the immunoprecipitation lanes in Fig. 4c), and that the input lanes ('Total Casp12') are duplicated. Similarly, the anti-tubulin control right lanes of Fig. 4b are duplicates of the left lanes. We are unable to reconcile how these images were incorrectly assembled despite diligent efforts to do so. Figure 1 of this Corrigendum shows a correct Fig. 4c, representing a new and independent experiment, to replace Fig. 4c of our Letter. The interpretation of the data and the conclusions are unaffected; namely, that caspase-12 forms a complex with and co-immunoprecipitates with caspase-1 when co-expressed by transfection into human (HEK293T) cells. There is also some association of caspase-12 with caspase-5 (more so than previously described) but very little with caspase-9. The experiment is robust and has been repeated a total of four times (twice each by two workers, M.S. and Claudia Champagne). We have also re-probed the original western blot in Fig. 4b with anti-tubulin and provide a replacement panel for that loading control in Fig. 1 of this Corrigendum. Our conclusions remain unaltered and the original legend for Fig. 4 also remains correct.

For clarity, more details for the legend to Fig. 2a are also provided. It should read: "…LacZ-neo cassette in the Casp12 targeted allele; β-galactosidase appears red). Panel I shows an intestine section from a wild-type mouse that was stained for β-galactosidase as a negative control. Panels II and III show intestine sections. Panel III is an enlarged view from Panel II. Panels IV–VII show splenic sections. Panel VI is another region of the section shown in Panel V.". The Supplementary Information to this Corrigendum shows the data used to generate Fig. 1 of this Corrigendum and Fig. 2a of our Letter.

Supplementary Information is available in the online version of this Corrigendum.

Correspondence should be addressed to D.R.G. (douglas.green@stjude.org).

1. Roy, S. *et al.* Confinement of caspase-12 proteolytic activity to autoprocessing. *Proc. Natl Acad. Sci. USA* **105**, 4133–4138 (2008); correction **110**, 4852 (2013).
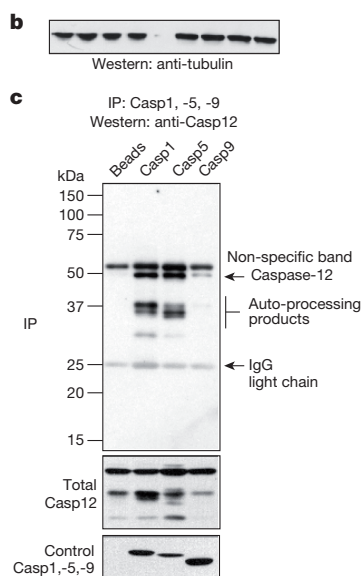
Figure 1 | This figure shows a correct Fig. 4c and Fig. 4b for our Letter.

CORRECTIONS & AMENDMENTS

Retraction: CLASP–mediated cortical microtubule organization guides PIN polarization axis

Klementina Kakar, Hongtao Zhang, Ben Scheres & Pankaj Dhonukshe

After re-examination of this Letter, concerns with some of the reported data were raised. It was found that two confocal images were near-identical in panels of Figure 3 and two confocal images were re-used in panels of Figure 4, and that some gel images were inappropriately generated by cutting and pasting of non-adjacent bands. Therefore, we feel that the most responsible action is to retract the paper. We sincerely apologize for any adverse consequences that may have resulted from the paper's publication.

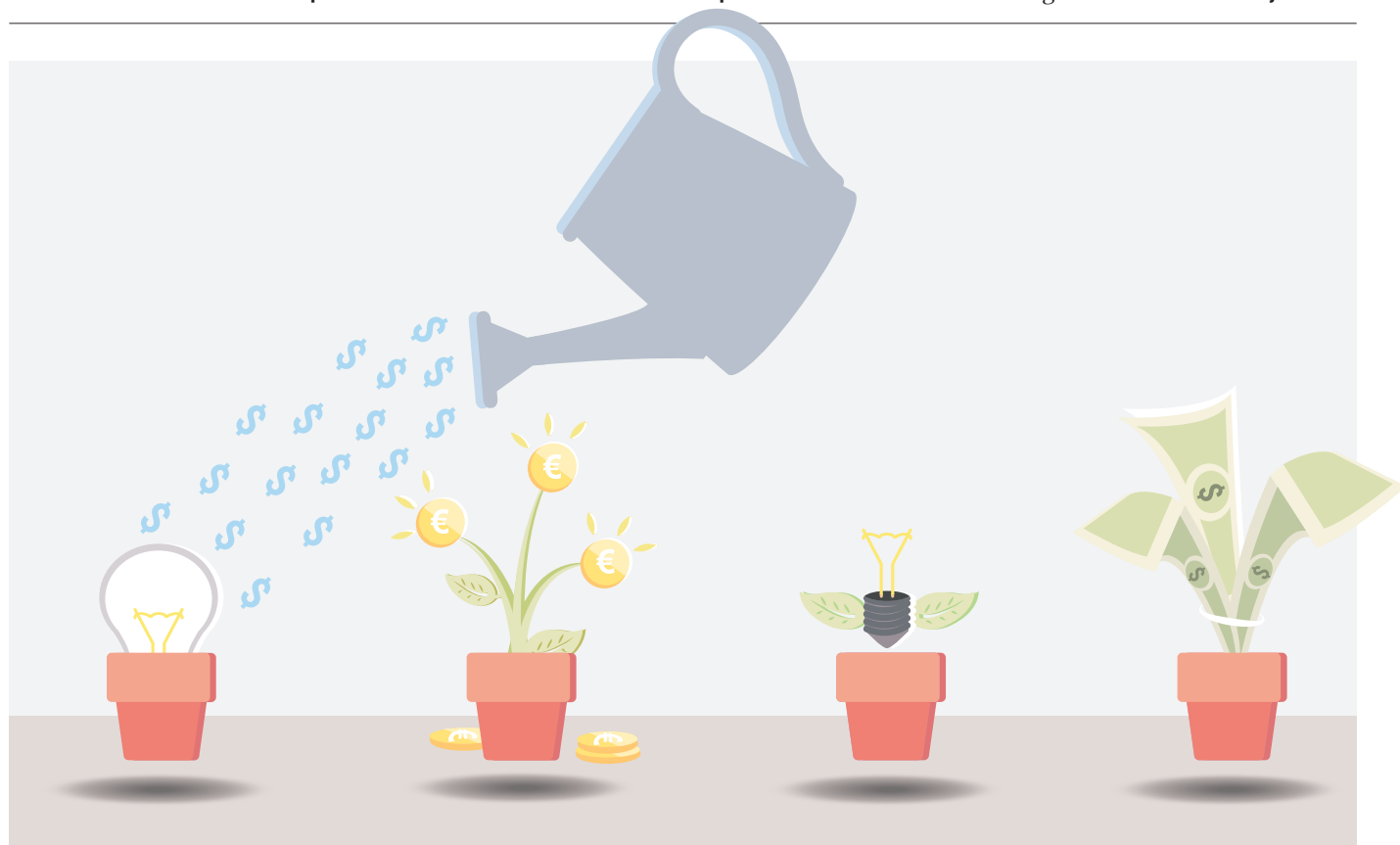**Correspondence** should be addressed to P.D. (p.b.dhonukshe@gmail.com).

CAREERS

COLUMN
A guide for the innovator

Researchers with product-worthy ideas can follow various, often under-appreciated, paths towards commercialization, says **Peter Fiske**.

Scientists often conduct research to advance innovations that they hope will improve lives. Indeed, for many it is a chief attraction of the scientific career.

But most are still unsure of how to recognize the commercial potential of their discoveries, and how to translate that into products or services. To the inexperienced or uninitiated, successful commercialization, when it happens, may seem to be a matter of chance — but that is not so. Although university technology-transfer offices often play a crucial part in helping to license inventions, many scientists do not realize the importance of their role in establishing early commercial validation for a new technology.

Historically, scientists have made their discoveries — and any commercial potential — known to the larger world mostly through scholarly publications. But academic literature, which is often behind paywalls and pricey to access, is aimed mainly at other academics. It is rarely consulted by industry. And even the most significant technological advance can become obscured by the detailed, turgid exposition that is the hallmark of the typical research article.

Researchers who think that their discoveries have commercial potential need to communicate this information through channels that industry accesses regularly. When publishing a scholarly article in a peer-reviewed journal, for example, it makes sense to notify the editors of the work's potential commercial impact. The journal might have some way of highlighting this facet of the work — be it through an editorial, an analysis or some other means. A number of journals now circulate brief synopses of their articles to a broader audience. Journal policies and approaches will vary, but scientists with entrepreneurial ambitions should not assume that the commercial aspects of their research will be noted unless specifically pointed out to editors.

There are other ways to promote a discovery's commercial potential. Some industry trade ▶

magazines welcome short, less-academic articles describing the innovation and its potential for industry. And many research institutions and universities have public-affairs departments that specialize in broadcasting newsworthy scientific discoveries, particularly ones with business potential. A scientist with an interesting invention can also reach out to specific industry communities that cluster around science blogs and other social-media outlets. LinkedIn, for example, has groups for 'oncology research', 'advanced biofuels' and many others.

Alternatively, researchers themselves can strike out as entrepreneurs. Creating, or 'spinning out', for-profit companies that are based on technologies developed in academic institutions is an increasingly crucial step in successful commercialization. Although the process of licensing remains alive and well at leading institutions, many tech-transfer professionals recognize that licensing alone is typically not enough to carry a new technology across the 'valley of death' — that difficult development stage in which research becomes more applied and costs escalate — and into commercial adoption. Furthermore, licensing to an outside party leads to a separation of the inventor from the invention, which can slow the development of a technology. After all, the researcher who first made the breakthrough often has the best insight into how to further develop the invention and resolve the remaining barriers to its commercial viability.

Academic start-ups seem to follow several paths (see go.nature.com/rqv7mc). Some researchers first explicitly market the innovation through their tech-transfer office to industry and the investment community. Many of the leading university tech-transfer offices maintain a network of industry contacts, ranging from Exxon to Genentech. An investor then comes forward, expressing interest in funding the company.

Some academic entrepreneurs have been able to capture, in essence, the best of both worlds by cultivating industry sponsorship for their product while the technology (and the research team) remains with the academic institution. In other words, they need not 'hand off' their discovery to the entrepreneurial sector. In some cases, industry partners specifically fund academic projects with the intent of developing a core set of technologies (and technologists) for later ingestion into the business. In many cases, the graduate students and postdocs who work on the projects are the first to be offered jobs with the industry partner. Investigators should, of course, tread carefully when

accepting such money, acknowledging and understanding any conditions of its acceptance (see *Nature* **453,** 1138–1139; 2008).

Sometimes, research projects can turn into commercial ventures with little outside assistance. In the course of their work, researchers or members of their team may produce a prototype of some sort — a new chemical, say, or a biological reagent. Investigators in academia or industry hear about that reagent and start to request it. The investigator starts charging a fee. The 'customers' get good results, demand the reagent in increasing quantities, perhaps become willing to pay a bit more and, voila, a business is up and running.

> *"Some institutions fund 'entrepreneur-in-residence' positions to specifically facilitate the creation of ventures based on scientific research."*

And then there is the explicit activity of entrepreneurs within the institute. Universities with business schools often have a cadre of students who trawl for technologies that have commercial potential in their institution's laboratories. These 'intrapreneurs' have an advantage because they are the first to learn the details of a discovery or innovation. Some institutions (such as the Lawrence Berkeley National Laboratory in California and the Massachusetts Institute of Technology in Cambridge) fund 'entrepreneur-in-residence' positions to specifically facilitate the creation of ventures that are based on scientific research. Getting to know your intrapreneurs can provide valuable insight not only into the process of technology commercialization but also into steering early-stage research in a commercial direction.

Academic and research institutions will continue to develop and refine the paths towards commercializing technological innovations. But the core of any successful effort is the researchers themselves. By understanding the pathways that can connect their discoveries to successful technological spin-offs, young scientists can maximize the chances that their research will have commercial value and impact. Big breakthroughs with commercial applications cannot necessarily be anticipated, but as the chemist Louis Pasteur noted, "fortune favours the prepared mind". ∎

ADAPTED FROM: PORNPHOL MANTHAM/SHUTTERSTOCK

Peter Fiske *is chief executive of PAX Water Technologies in Richmond, California, and author of* Put Your Science to Work *(American Geophysical Union, 2000).*

TURNING POINT
Tracey Holloway

More than ten years ago, Tracey Holloway and her colleagues started what has become the 1,700-member Earth Science Women's Network. Now an air-quality and energy researcher at the University of Wisconsin–Madison, Holloway is helping to turn the network into a non-profit organization.

Did you have early intentions to pursue a science career?

No. In fact, a school friend who believed that women self-select out of science made me realize that I may have been doing that. I kept an open mind and found that I liked chemistry and maths as an undergraduate more than I had as a high-school student. After my first year, I spent a summer internship at the Johnson Space Center in Houston, Texas, and realized that I could use fluid dynamics to look at hurricanes, global warming and air pollution. That put me on track for a graduate degree in atmospheric sciences at Princeton University in New Jersey.

Describe the gender dynamics during your graduate experience.

When I started, there were only ten graduate students across all years. There were a couple of women in my first year, but for a while after that I was the only woman. The faculty were all men.

Did your PhD go smoothly?

I didn't find a project I felt passionately about in the first year and wondered whether I had made the wrong choice. In fact, I asked the department chair for a leave of absence. He told me I had to send a letter to each faculty member explaining what I'd do during the year off, and the response was surprising. They were very supportive, and several talked to me about my interests in policy and law. At that time, Princeton had started a programme through the Woodrow Wilson School of Public and International Affairs that allowed science students to do a policy-relevant project, so they encouraged me to go down that route.

When did you start the Earth Science Women's Network?

During graduate school, several women colleagues and I hosted a session at the American Geophysical Union (AGU) in Washington DC on policy-relevant science. After the session we talked about our careers, spouses and work–life balance, and we realized that we wanted to stay in touch beyond these meetings. It started out with six of us on an e-mail list, but it quickly grew into what we described as a peer-mentoring network of 25 members, so we set up an electronic mailing list. Over time, the network received financial support from the AGU, the US National Center for Atmospheric Research in Boulder, Colorado, and the US National Science Foundation in Arlington, Virginia.

Was there any backlash against forming a women-focused network?

No, but we were always very aware that it could be perceived as sexist. That's one reason why, early on, we split off a jobs network to serve the entire early-career community. But we built this network to have conversations that weren't happening elsewhere, and it clearly served that role.

Why the move to become a non-profit?

Until last year, we had support from universities, but the funding expired. We were trying to figure out how to move forward, and becoming a non-profit seemed like the best move to cover our expenses. But it costs money to start a non-profit, so we launched a crowd-funding campaign on Crowdtilt in January. It's been really inspiring to see the support that came back. We've raised just shy of US$14,000 from roughly 300 contributors.

What impact has this network had on your career?

I think it's made my career experience more satisfying. The women in this group are my closest friends because I'm able to connect with them in all facets of my life. We talk about research one minute and our kids the next. I've found that I've got the best advice from people at or near my career stage. ∎

INTERVIEW BY VIRGINIA GEWIN

Dwindling tenure posts

The proportion of non-tenure-track and non-tenured faculty posts continues to rise across all US institutions, finds a report by the American Association of University Professors (AAUP) in Washington DC. *Losing Focus: The Annual Report on the Economic Status of the Profession, 2013–14* surveyed 1,159 public and private US institutions and found that the overall proportion of assistant professors in non-tenure-track posts was 23.4% for 2013–14, compared with 20.8% in 2010–11. Dwindling tenured and tenure-track posts threaten the ability of scientists to conduct research without interference from funders or administrators, says John Curtis, the report's lead author and director of research and public policy for the AAUP.

Student debt rising

Education debt owed by US holders of master's degrees in science is snowballing, says a report from the New America Foundation, a non-profit public-policy group in Washington DC. Half of the graduates owed at least US$36,000 in 2012, up 28% from 2008 levels. One-tenth owed $92,126 or more, up 36% from 2008. The proportion graduating with debt rose by almost 8%, to 53.6%, between those years. *The Graduate Student Debt Review* examined data that the US Department of Education collects every four years. Report author Jason Delisle speculates that the increasing indebtedness will eventually spur government-led loan-forgiveness programmes.

Grant obstacles

Tough grant-proposal deadlines can have severe impacts on researchers, according to an Australian study. The authors surveyed 215 academic researchers across all career stages in Australia who prepare annual applications for project grants from the National Health and Medical Research Council (D. L. Herbert *et al. Br. Med. J.* **4,** e004462; 2014). Most respondents reported that the 2–3-month application process each year takes top priority, superseding research and publishing, their health and their personal responsibilities. Almost all said that they would support adding more grant cycles and deadlines to lessen the burden for applicants.

HOW KAMERON LAYAS RODE OUT THE CRASH

Shameful behaviour.

BY RAHUL KANAKIA

For about 30 seconds back in 2031, Kameron Layas was the most famous shame engineer in the world.

His breakthrough came when he bought 30 years of archived video from the surveillance cams of a very large supermarket chain.

Thirty years' worth of frazzled mums and dads pleading with hyperactive kids who won't stop yapping in their ear about some $7 piece of neon-coloured candy gunk — pleading for just one minute — no, one second — of quiet, so that they can just try to remember if they have everything on the list …

Thirty years of parents reaching their limit — just once … once in a lifetime — and buying that moment of quiet by slapping their kid upside the head.

And then, in the summer of 2031, 30 years' worth of parents — many of them with kids who'd long ago matured into a healthy Oedipal hate — got official-looking letters containing pictures of the long-denied incident … and those letters came with fines. The mums and dads paid up. You bet they did.

It was the mother of all breakthroughs. Within six months, the shame engineers had poured into parenting.

If you fed your kid too much mercury-tuna, or didn't strap them into a child-seat, or let them ride without a helmet, then those ubiquitous silently recording cameras got it all, and you probably got a notice. Of course, the bonanza didn't last long. It never does. By the time Kameron made the cover of *Time*, he was almost bankrupt, and the big corporations had already moved into the parental-shame business. They started putting microchips in all their products — cars, TVs, guns, everything — so that they could record whether parents were exposing their kids to danger. How could the smaller players compete with that?

That's when Kameron began his long, slow descent. However, he refused to feel ashamed of his decline. He spat on the sidewalk and he shat in the street and he kicked at cars and pried the microchips out of things, and the fines kept piling up, oh they kept piling up … Kameron owed hundreds of thousands — millions! — of dollars in fines that there was no way on Earth he was ever going to pay.

And then Kameron overdosed.

Kameron's best friend in the world was Doctor Exponential. They worked the parks together, collecting bottles. At night they pushed their shopping carts back to the overpass and listened to the music of the freeways. Sometimes they'd pool their cash and score a little horse from the guy at the corner.

But nothing in their partnership really covered what to do during an overdose. Kameron was huge: he weighed more than 300 pounds. The Doctor was small and skinny. How was he supposed to drag Kam to the hospital?

Besides, the Doctor could see the three cameras watching them. They knew Kam was dying. If they wanted, they could save him, but the Doctor knew that the cameras didn't give a damn about people like Kam.

But this time he was wrong.

As soon as the cameras identified Kameron Layas, a computer alerted Oakland's private-ambulance service. The little self-driving rescuecar zipped over and scooped Kameron up while Doctor Exponential was still looking around for help. The Doctor turned just in time to see the ambulance carry his friend away.

The Doctor wasn't there when Kameron woke up. The Doctor was at the library, researching. Kameron was discharged with a full belly and a $24,000 bill. It took him the best part of a day to pick his way back to the underpass.

The two of them clasped wrists when they saw each other. They sat down, leaning against the pylon, and the Doctor covered Kameron with a blanket.

"Hey man," said the Doctor. "You know you're worth more alive than dead?"

The two of them sat there like that for ten minutes. Kameron glanced at the dealer on the corner: the dude had a gun tucked just under his shirt.

"What'd you mean by that?" Kameron asked.

"You've got total fines of like … ten million dollars."

"Yeah?"

Back in the day, the Doctor had been a real doctor: a doctor of maths. He said: "Every time you get a fine, a computer thinks you owe some company some money."

"So what?"

The Doctor grunted. How had Kameron managed to do the things he did? Kam never understood anything.

"If someone owes ten mil to a company, then that company, in its account books, counts that debt as an asset: they start acting ten mil richer. Then they slice that debt up and use it to back securities that they sell to other people. Over time, it all gets multiplied, so if you'd died, a hundred mil worth of securities would've become worthless. That's why you're still alive."

"Yeah?" Kameron said. "Alright …"

Kameron looked at the nearest camera. And then he smiled. It was pointed straight at him.

He got up and walked over to the dealer. "Hey, give me all your vials," he said.

The dealer said: "What? Get out of here, skag."

Kameron stepped up, reached into the guy's pockets, and pulled out two handfuls of glass vials. The dealer said: "What the …?" and pulled out his gun. But when he pulled the trigger, it blinked red and did nothing.
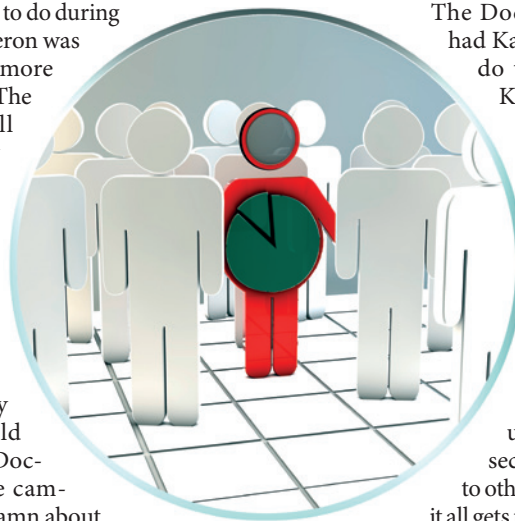
"Nuh uh," Kameron said. "The computer ain't gonna let you mess with me."

Kameron pulled the gun out of his hand, then he fired it once, up into the air, and the dealer started running away.

Kameron stepped into the road without looking, but the self-driving cars did rapid swerves right around him. He walked back and dropped all the vials onto Doctor Exponential's lap. "Take them," Kameron said. "I don't need them any more."

Then he sat down next to his friend, lay the gun in his lap, and covered it with a corner of the blanket. ∎

Rahul Kanakia *is a science-fiction writer based in Baltimore, Maryland.*